2011-03

Aviation selection test battery component predictiveness of primary flight training outcomes among diverse groups

Denton, Tremain L.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AVIATION SELECTION TEST BATTERY COMPONENT PREDICTIVENESS OF PRIMARY FLIGHT TRAINING OUTCOMES AMONG DIVERSE GROUPS

by

Ramon A. Lopez
Tremain L. Denton

March 2011

Thesis Advisor:	John Schmidt
Second Reader:	Scott Nestler

THIS PAGE INTENTIONALLY LEFT BLANK

<table>
<tr><td colspan="3">REPORT DOCUMENTATION PAGE</td><td>Form Approved OMB No. 0704-0188</td></tr>
<tr><td colspan="4">Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.</td></tr>
<tr><td>1. AGENCY USE ONLY (Leave blank)</td><td colspan="2">2. REPORT DATE
March 2011</td><td>3. REPORT TYPE AND DATES COVERED
Master’s Thesis</td></tr>
<tr><td colspan="3">4. TITLE AND SUBTITLE
Aviation Selection Test Battery Component Predictiveness of Primary Flight Training Outcomes Among Diverse Groups</td><td rowspan="2">5. FUNDING NUMBERS</td></tr>
<tr><td colspan="3">6. AUTHOR(S) Ramon Lopez, Tremain Denton</td></tr>
<tr><td colspan="3">7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000</td><td>8. PERFORMING ORGANIZATION REPORT NUMBER</td></tr>
<tr><td colspan="3">9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A</td><td>10. SPONSORING/MONITORING AGENCY REPORT NUMBER</td></tr>
<tr><td colspan="4">11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number: N/A</td></tr>
<tr><td colspan="3">12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited</td><td>12b. DISTRIBUTION CODE
A</td></tr>
<tr><td colspan="4">13. ABSTRACT (maximum 200 words)
The Aviation Selection Test Battery (ASTB) has been the qualifying benchmark for the Naval Aviation since World War II. While it is necessary that test scores effectively select the candidates with the greatest chance for success, the ASTB strides toward increasing diversity while maintaining low attrition. Using archived Student Naval Aviator and Student Naval Flight Officer ASTB subtest scores and Primary Flight Training (PFT) records, this study examined the ASTB’s predictive ability with respect performance in PFT. Specifically the study consists of two analyses: 1) determine how well the ASTB could predict majority and minority group performance in primary flight training; and 2) determine how well the ASTB could predict success in each training phase and for the entire sample and select groups. The linear regression analysis successfully fit a significant model for the entire sample and Caucasians, but was unable to produce a significant model for African Americans or Hispanics, as there was insufficient data available for either group. The model, when fitted to the entire dataset, with race as an independent variable, yielded a result where all independent variables were significant. The results from the logistic regression models showed there was evidence that four of the ASTB subtests were significant and positive predictors for the entire sample and Caucasians; but was unable to produce a significant model for African Americans or Hispanics. It is apparent that the small data set for minorities limited this study. Efforts to collect data from personnel records should be conducted to obtain all scores from flight training, so that these groups can be further investigated.</td></tr>
<tr><td colspan="3" rowspan="2">14. SUBJECT TERMS
Selection, Aviation Candidate Selection, Training, Diversity, Diversity in Flight Training, ASTB, Military Selection</td><td>15. NUMBER OF PAGES
97</td></tr>
<tr><td>16. PRICE CODE</td></tr>
<tr><td>17. SECURITY CLASSIFICATION OF REPORT
Unclassified</td><td>18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified</td><td>19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified</td><td>20. LIMITATION OF ABSTRACT
UU</td></tr>
</table>

NSN 7540-01-280-5500

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

AVIATION SELECTION TEST BATTERY COMPONENT PREDICTIVENESS OF PRIMARY FLIGHT TRAINING OUTCOMES AMONG DIVERSE GROUPS

Ramon A. Lopez
Lieutenant, United States Navy
B.S., Southern Illinois University Carbondale, 2000

Tremain L. Denton
Lieutenant, United States Navy
B.S., Excelsior College, 2001

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN HUMAN SYSTEMS INTEGRATION

from the

NAVAL POSTGRADUATE SCHOOL
March 2011

Authors: Ramon A. Lopez
Tremain L. Denton

Approved by: John Schmidt
Thesis Advisor

Scott Nestler
Second Reader

Robert Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Aviation Selection Test Battery (ASTB) has been the qualifying benchmark for the Naval Aviation since World War II. While it is necessary that test scores effectively select the candidates with the greatest chance for success, the ASTB strides toward increasing diversity while maintaining low attrition. Using archived Student Naval Aviator and Student Naval Flight Officer ASTB subtest scores and Primary Flight Training (PFT) records, this study examined the ASTB's predictive ability with respect performance in PFT. Specifically the study consists of two analyses: 1) determine how well the ASTB could predict majority and minority group performance in primary flight training; and 2) determine how well the ASTB could predict success in each training phase and for the entire sample and select groups. The linear regression analysis successfully fit a significant model for the entire sample and Caucasians, but was unable to produce a significant model for African Americans or Hispanics, as there was insufficient data available for either group. The model, when fitted to the entire dataset, with race as an independent variable, yielded a result where all independent variables were significant. The results from the logistic regression models showed there was evidence that four of the ASTB subtests were significant and positive predictors for the entire sample and Caucasians; but was unable to produce a significant model for African Americans or Hispanics. It is apparent that the small data set for minorities limited this study. Efforts to collect data from personnel records should be conducted to obtain all scores from flight training, so that these groups can be further investigated.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AFOQT	Air Force Officer Qualifying Test
ANIT	Aviation/Nautical Information Test
API	Aviation Preflight Indoctrination
AQR	Academic Qualification Rating
ASTB	Aviation Selection Test Battery
ASVAB	Armed Services Vocational Aptitude Battery
BI	Biographical Inventory
CNATRAINST	Chief of Naval Aviation Training instruction
DOR	Drop on Request
FOFAR	Flight Officer Flight Aptitude Rating
IFS	Introductory Flight Screening
IV	Independent Variable
KSA	Knowledge, Skills, and Abilities
MCT	Mechanical Comprehension Test
MST	Math Skills Test
NAMI	Naval Aerospace Medical Institute
NASC	Naval Aviation Schools Command
NPQ	Not Physically Qualified
NROTC	Naval Reserve Officer's Corps
NSS	Navy Standard Score
OCS	Officer Candidate School
OOM	Order of Merit
PBMB	Performance Based Measures Battery
PFAR	Pilot Flight Aptitude Rating
PFS	Flight School/Primary Flight Training
RCT	Reading Comprehension Test
SAT	Scholastic Aptitude Test / Spatial Apperception Test
SNA	Student Naval Aviators
SNFO	Student Naval Flight Officer
USMC	United States Naval Academy

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Naval Aviation has targets in place with respect to the number of Student Naval Aviator (SNA) and Student Naval Flight Officer (SNFO) candidates that can be accepted into training in a given fiscal year. While it is necessary that test scores effectively select the candidates with the greatest chance for success, the Aviation Selection Test Battery (ASTB) strides towards increasing diversity while maintaining low attrition rate must continue. This research provides the Naval Aviation Medical Institute (NAMI) an analysis of the predictive ability of the ASTB subtests for success in SNA and SNFO primary flight training pipelines. It also provides insight into the subtests' varying predictive reliability among diverse racial/ethnic groups in the aviation training pipeline. The overall predictive ability is then examined for each phase of the primary flight training.

The hypotheses tested in this study are as follows:

$H1_O$: There is no difference between the predictive ability of the ASTB in minority and majority SNAs and SNFOs primary flight performance.

$H2_O$: There is no difference in predictive ability of the ASTB for the overall success rate at the end of PFS between minority and majority SNAs and SNFOs.

$H3_O$: There is no difference in predictive ability of the ASTB for success in the earlier phases of flight training (Aviation Preflight Indoctrination and Introductory Flight Screening) between minority and majority SNAs and SNFOs.

The data set analyzed was comprised of Naval officers, sourced from the Naval Academy, Naval Reserve Officers Training Corps, or Officer Candidate School who entered service from fiscal years 2002 to 2010. Initially, the dataset was separated into majority and minority categories by race/ethnicity. The majority category was composed of those who reported themselves as Caucasian, while the minority group was composed of those who self reported as African American or Hispanic. To complete the analysis for the global model

used in the first hypothesis, the minority group was sectioned into African Americans and Hispanics. Gender was not addressed in this study, as males constituted the majority of the available sample.

The linear regression analysis successfully fit a significant model for the entire sample and Caucasians, but was unable to produce a significant model for African Americans or Hispanics, as there was insufficient data available for either group. The model, when fitted to the entire dataset, with race as an independent variable, yielded a result where all independent variables were significant.

The results from the logistic regression models showed there was evidence that four of the ASTB subtests were significant and positive predictors for the entire sample and Caucasians; however, these models only explained a small portion of the total variance. It was apparent that the small data set for minorities limited this study. Efforts to collect data from personnel records should be conducted to obtain all scores from flight training, so that these groups can be further investigated.

ACKNOWLEDGMENTS

From Lieutenant Ramon A. Lopez:

I would like to acknowledge those who made this thesis work possible:

First and foremost, I would like to thank my wife, Amanda. I am eternally grateful for your patience, love and understanding. Thank you for being by my side through it all.

I would like to express my sincere gratitude to my thesis advisor, USN CAPT John Schmidt and Second Reader USA LTC Scott Nestler.

Thanks for your support, instruction and supervision, without your guidance, this thesis would not be possible.

To USN LCDR Hank Philips, NOMI and USN LCDR Chris Foster, CNATRA, thank you for your work and dedication into this study.

To my Human Systems Integration classmates, thank you for not letting me falter in times of need; you all pitched in on this.

To the Human Systems Integration staff and faculty, thank you for assisting me in completion of my schooling and this thesis.

To my mother, my sincerest thanks for your guidance and encouragement in all that I do. Your strong heart helped me through the most difficult time of my life.

Lastly, I would like to dedicate this to my father, who has shown me how hard work, dedication, and persistence can overcome all. Rest in peace.

From Lieutenant Tremain L. Denton:

I would like to thank my loving wife of thirteen years for believing in me in my moments of self-doubt, and helping me remain centered when all seemed to be collapsing around me. You are an angel.

I would like to thank my thesis advisor, CAPT John Schmidt, for taking a chance on me. Without his amazing patience, guidance, and vision—I don't know what I would have done. Thank you to LTC Scott Nestler for helping to make the numbers "tell a story."

I would like to thank the Human Systems Integration faculty and staff for their instruction help and support.

Last, but certainly not least, I would like to thank my parents for never accepting less than my best, I owe all that I am to you.

I. INTRODUCTION

A. OVERVIEW

Personnel selection in the American military has its early roots in World War I, when the U.S. Army first incorporated the use of aptitude tests to screen people for military service in addition to using entrance physicals (Yerkes, 1921). The Army Alpha and Beta tests provided leaders a measure of individual ability in making personnel assignments (ASVAB, 2011). These tests were incorporated due to the volume and variability of the volunteers and draftees being inducted. The military use of aircraft led to an emerging need to systematically train pilots for the first time, and consequently identify candidates for flight training (Pohlman, & Fletcher, 1999). Because those vying to fly came into military service with little or no flight experience, diverse educational backgrounds, and varied physical attributes, a clear need for a selection process arose (Carretta & Ree, 2003). Consequentially, screening tests assessing reaction time, cognitive ability, equilibrium, and emotional stability was established (Henmon, 1919).

The use and importance of effective selection was magnified in World War II, given the scope and duration of the conflict (Flanagan, 1942). This was particularly true for aviation, where high demand for aviators was coupled with high attrition rates (>50%), training costs, and accident rates (Burke & Hunter, 1995). Initial selection processes consisted of physical assessments, biographical interviews, and general intelligence tests (Jenkins, 1946). These screening mechanisms were somewhat helpful in lowering attrition, but accident rates remained relatively high. Not surprisingly, there were more losses experienced in training than there were in combat (USAAF, 1945). As aircraft technology continued to advance during the war it led to greater maneuverability, faster speeds, and higher service ceilings and greater demands on aircrew (Hilton & Dolgin, 1991). This made the need for developing an effective selection process that much greater.

The U.S. military pressed with the development of selection tests to reduce attrition and increase safety. The Army Air Forces developed the Aviation Cadet Qualifying Examination, which used an array of paper and pencil, motion picture, and apparatus tests designed to assess leading attrition factors, including: intelligence, judgment, alertness, observation, decision speed, reaction time, coordination, emotional control, motivation, and ability to divide attention (Flanagan, 1942). Naval Aviation in contrast continued to rely on physical screening and refined biographical inventories, intelligence tests, aptitude tests, and targeted interviews (Fiske, 1947). Despite the variation in approach both processes served to reduce training attrition (Pohlman, & Fletcher, 1999).

Today, the need to provide for an effective aircrew selection process is still great given the cost to train the average military aviator is nearly a million dollars (GAO, 1999). The U.S. Air Force has maintained its own unique selection process; although it was modified and revalidated over the years, it still reflects the unique characteristics stemming from its initial development during World War II (Burke & Hunter, 1995). The U.S. Air Force administers the Air Force Officer Qualifying Tests (AFOQT), which is a standardized paper-and-pencil instrument with 12 subtests tapping into: verbal analogies, arithmetic reasoning, word knowledge, instrument comprehension, block counting, table reading, aviation information, general science, rotated blocks, hidden figures, and self-description inventory (USAF ROTC, 2009). It is now complemented with the use of an apparatus test called the Basic Attributes Test, which assesses psychomotor, cognitive, and personality measurements (Carretta & Ree, 2003).

The U.S. Army, with greater utilization of helicopters during the Vietnam era, developed the Flight Aptitude Selection Test. It consists of seven subtests: background information, instrument comprehension, complex movements, helicopter knowledge test, cyclic orientation test, mechanical functions test, and self-description (Dept. of Army, 2005). It was not an intelligence test, but rather one of aptitudes and characteristics predictive of Army helicopter flight training success. This test has also evolved since its initial development, but retained its unique focus on the selection rotary-wing aviators (Wiener, 2005).

Naval Aviation continued to modify and revalidate its selection process (Berkshire, 1967). Now called the Aviation Selection Test Battery (ASTB), it retained the use of cognitive tests to assessing intelligence, ability, and aptitude, but recently dropped biographical information (NAMI, 2011). Presently, the ASTB enjoys the highest predictive validity among the various service selection tests for primary flight training performance (NAMI, 2011). The ASTB is now delivered on-line (Olde, Olsen & Phillips, 2007), and is augmented with the development of an apparatus test using performance-based measures to delve into current factors driving attrition. These measures are tied to task saturation, task fixation, and an inability to switch between tasks (Olde, Olsen & Walker, 2007).

Each service-specific test tends to be unique in content and scope, yet all have measures of cognitive ability in common. Such cognitive-based tests in recent years have been seen as problematic because they generally lower pass rates among minority candidates and lower predictive ability for success in training (Outtz, 2002). Unfortunately, military selection tests have not been immune to this problem; consequentially, much attention is paid to ensure measures are taken to eliminate potential sources of test bias and provide a level playing field (Caretta, 1997). While the main mission of selection tests in military aviation is to minimize attrition and associated costs, they must also provide fairness if a level of diversity is to be achieved.

B. OBJECTIVE

This thesis has three primary goals 1) assess the effectiveness of the ASTB in predicting Student Naval Aviator (SNA) and Student Naval Flight Officer (SNFO) success in flight training, 2) determine if the ASTB is equally effective in predicting minority/majority SNA and SNFO success in flight training, and 3) assess the effectiveness of the ASTB in predicting SNA and SNFO success by flight training phase. The Navy and Marine Corps test over 10,000 aviation applicants annually of which 15% are ultimately selected for flight training (Williams, Albert, & Blower, 2000). According the GAO (1999) the cost for training an aviator is about $1 million, and with the high volume of SNAs and SNFO in flight training total annual costs tops $1.5 billion. Clearly, reduced attrition through effective selection can equate to significant savings, and the perspective gained from this effort could lead to further improvements. Finally, this study will provide greater insight with respect to differences in utility for minority and majority candidates.

C. AVIATION SELECTION TEST BATTERY

The Flight Aptitude Rating, introduced in 1942, was developed by the Naval Aerospace Medical Institute (NAMI) in Pensacola, Florida (NAMI, 1991).Today the Naval Services including the U.S. Navy, U.S. Marine Corps, and U.S. Coast Guard, all use a current, expanded version for pilot selection known as the ASTB. It is the sole testing tool for making selection determinations for potential aviator applicants (NAMI, 2011). The most current version of the ASTB released in 2004, has three different forms and is composed of five subtests: Math Skills Test (MST), Spatial Apperception Test (SAT), Mechanical Comprehension Test (MCT), Aviation/Nautical Information Test (ANIT), and Reading Comprehension Test (RCT) (NAMI, 2011). Even though it originated as a paper-and-pencil test (Williams et al., 2000), it is administered now primarily on the Internet (NAMI, 2011).

The ASTB plays an early role in the selection process, as it acts as a filter in narrowing down field of applicants for training (Williams, et al., 2000). The current ASTB was constructed and validated to predict both performance and attrition through the primary phases of SNAs and SNFOs training and saves Naval Aviation over $30 million annually (NAMI, 2011). Applicants receive scores derived from combinations of the ASTB subtests that are used for the selection; they are the Academic Qualification Rating (AQR), Pilot Flight Aptitude Rating (PFAR), and Flight Officer Aptitude Rating (FOFAR). The AQR predicts academic performance in aviation preflight indoctrination and the ground school phase of primary flight training (NAMI, 2011). The PFAR and FOFAR predict flight training performance for SNAs and SNFOs respectively. Each ASTB component scores is reported in stanines, normalized standard scores with a range of 1 to 9, a mean of 5, and a standard deviation of 2. Similar to percentile ranks, they are status scores within a particular norm group. For the purpose of this study, the focus is on individual subtests results, rather than the derived composite scores in stanine format.

The ASTB is open to any physically qualified candidate with the desire to be a Naval Aviator. As previously mentioned, the ASTB test is administered to nearly 10,000 applicants each year, of which only 15% are selected for flight training (Williams et al., 2000). Data gained from NAMI depicts that 90% of the applicants who successfully completed the selection process were Caucasians. This would suggest that, of the 15% selected for flight training, 1,350 SNAs out of 1,500 are Caucasian. According to a Department of the Navy (2010) diversity study, targeted officer accessions is a key element for achieving a more diverse workforce. The ASTB is not a mechanism designed to produce a diverse workforce; it is a tool for selection, determined to select the right candidate for flight training and reduce attrition. It has, however, been extensively validated for effectively predicting success in training and against statistical bias, in terms of race, ethnicity, and gender (Dean, 1996).

D. RELEVANT HUMAN SYSTEM INTEGRATION (HSI) DOMAINS

The present study examines the ability to predict performance and success in flight training, based on the scores achieved on the components of the ASTB. It taps primarily into three of the eight HSI domains: Manpower, Personnel, and Training. The following is a characterization of each domain and how it is combined into the present study.

Manpower: The number and composition of people, who operate, maintain, support, and provide training for a system (Booher, 2003). It provides insight into the number of candidates needed to be recruited and trained to meet demands for qualified aircrew. It is also tied to Naval Academy and NROTC graduates not pursuing a naval career if they attrite from flight training.

Personnel: The selection of individuals with appropriate knowledge, skills, and abilities required to perform as operators, maintainers, or support personnel (Booher, 2003). Selecting the right candidates would mean having a better chance of success in the flight program. An effective selection process would also provide fewer misses of capable diversity candidates for training.

Training: This involves the instruction, education, and training required to provide personnel with the knowledge, skills and abilities needed to operate and maintain systems (Booher, 2003). With an effective selection process and training program in place, there would be a greater chance of success and provide a reduction in attrition rates. Less flight program attrition would equate to a reduction in required training resources (aircraft, instructors, and simulator time) needed.

Collectively and separately, ensuring that manpower, personnel, and training, as the HSI domains are properly addressed in selection, will yield benefits in terms of enhanced productivity, cost reduction, force stability, and greater diversity.

E. ORGANIZATION

This thesis contains five chapters. Chapter I covered the challenge of aviation selection, objective for the study, and relevant domains of HSI involved. Chapter II reviews related literature of the selection process for potential pilot candidates. Chapter III details the data and methods employed. Chapter IV contains the study results. Chapter V presents the conclusions and recommendations. The appendices contain tables depicting components of the analysis summarized earlier in Chapter IV.

II. LITERATURE REVIEW

A. OVERVIEW

Given the need to effectively identify candidates for aviation training, a selection process is set in place to make appropriate candidate decisions so as to minimize training attrition, best utilize training resources, minimize cost, and promote safety. In order to comprehend how to achieve this end, it is important to understand the foundational elements of an effective selection process. The purpose of this chapter is to provide a background on the selection process, review representative literature that relates it to military aviation selection, and then specifically in Naval Aviation. The review examines literature covering selection effectiveness, specifically in terms of validity, reliability, and utility. It then touches upon some ethical and legal implications in selection, promoting diversity, and ensuring fairness in selection.

B. SELECTION

The personnel selection process is an early step in determining the best applicants for positions in given career fields (Aamodt, 2004). It is an important process in situations where there are more qualified individuals than open positions. The goal of selection is to capitalize on individual differences in order to identify candidates who possess a determined amount of particular characteristics judged important for job success (Cascio, 1998). Decisions in personnel selection focus on matching an individual to a position where their requisite knowledge, skills, and abilities (KSAs) meet or exceed the identified requirements of the position. To have an effective selection process there is a need to conduct a job analysis that identifies KSAs for a specific job and the level required for each one identified. Based on the KSAs developed, a selection criterion is established for a given position.

A typical selection process can consist of several inter-related activities including application, interviewing, and testing (Cascio, 1998). The following is a brief characterization of each component:

Application: an initial step that involves the applicant providing pertinent personal, education and training, and work experience information. Often applicants provide a resume with a personal history.

Interviewing: employers in face-to-face interaction verify and obtain information. It provides an impression of the applicant and their communication skills.

Testing: provides employers an assessment on an applicant's KSAs for potential job placement. Effectively used it can save training time, material, and resources as well as lead to greater job performance.

Employers also often have additional selection procedures for specific jobs that are tied to health, safety, and security requirements. Among them are physical exams, drug screening, and background checks (Aamodt, 2004).

C. EFFECTIVENESS AND FAIRNESS

Key elements to provide for effectiveness and fairness in selection are validity, reliability, utility, and fairness (Cascio, 1998). Each of these are essential to have a legal, ethical, and beneficial process in place for identifying suitable candidates for employment while safeguarding against bias which may hinder it. The following paragraphs characterize these critical components.

Validity is the most fundamental test issue and is the extent to which a procedure actually captures what it is designed to measure (Proctor & Van Zandt, 2008). In a selection process, validity is the degree to which a measure accurately predicts job performance (Aamodt, 2004). The classical validity approach to personnel selection places primary emphasis on measurement accuracy and predictive efficiency.

There are three major processes used to validate predictors they are: Construct, Criterion-Related, and Content Validity (Cascio, 1998; Aamodt, 2004). Construct validity is the most important form of test validity and refers to the extent to which the test measures what it purports to be measuring. Criterion-related validity refers to the extent to which the test predicts a criterion. If the criterion is measured about the same time as the test is administered the term “concurrent validity” is used, in contrast to “predictive validity,” which is used when a certain period has passed between testing and criterion measurement. Content validity is a third type of validity, concerning the extent to which the test items or questions are covering the relevant domain measured.

Reliability is defined the stability of a measurement over time (Proctor & Van Zandt, 2008). It is essential for a test to be considered reliable for it to also to be deemed valid (Aamodt, 2004). Test reliability can be significantly affected by interruptions, time of day, etc., therefore, making standardized administration a requirement.

Reliability in testing is primarily tied to that between administrations and across parallel forms (Cascio, 1998; Aamodt, 2004). Test-retest reliability evaluates reliability across time, in that performance at one point in time on a similar test short correlate with the second score achieved. Reliability is also assessed between two parallel forms of an instrument, where a high correlation is expected between the scores received on similar tests.

Utility refers to the overall usefulness of a personnel selection procedure (Cascio, 1998; Aamodt, 2004). The concept focuses on the accuracy of the predictor and the importance of personnel decisions, the costs and benefits of selection decisions in terms of errors made, and the expense of setting up and implementing selection procedures. It also encompasses the selection ratio, the ratio of the number of available openings to the total number of available applicants and the base rate of success, the proportion of people successfully placed in the available openings using the selection criteria.

Fairness is the premise of achieving equity with respect to selection processes (Aamodt, 2004). The degree to which an instrument achieves an acceptable level of fairness is dependent in part upon the composition of the pool from which candidates are to be selected, the range of performance levels present, and the appropriateness of the performance level deemed required. Investigations of unfair discrimination must consider job performance in addition to the predictor of performance. A selection measure cannot be said to discriminate unfairly if inferior predictor performance by a group also is associated with inferior job performance by the same group (Cascio, 1998).

D. SELECTION IN MILITARY AVIATION

As observed in the introduction, both the Army and Navy in World War I were using tests in conjunction with physical examinations and biographical interviews for pilot candidate selection. In World War II with increased need for effective, efficient selection processes couple with advances in flight system technology the tests moved from primarily general ability measures to more tailored measures of identified KSAs, to include reading comprehension, spatial orientation, and mechanical understanding. Over the decades that followed efforts to enhance selection have persisted to further minimize attrition and avoid associated sunk personnel and training costs (Burke & Hunter, 1995).

Today across the services similar processes for selecting candidate for flight training are employed. All require physical examinations, background checks, interviews, and drug screening as well as meeting a cut off score on a standardized selection test. In recent years with the advent of modern computer technology and the emergence of the internet, computer based testing has emerged. All three services to a varying extent have incorporated both in their selection processes. Leading the way in this application of technology was Naval Aviation with its development of the Automated Pilot Examination System, which supports ASTB administration, scoring, reporting, and archiving (Carretta & Ree, 2003).

The key elements for effectiveness and fairness in the selection process are no different when applying them to the field of military aviation selection. As mentioned earlier in this chapter, the selection process is to determine the best applicant for a position. For military aviation, the selection process begins with testing to use as a predictor for job performance. The validity of these tests is crucial in military aviation, especially when the cost to train an aviator is nearly $1 million dollars (GAO, 1999; Martinuseen & Hunter, (2010)); there is great emphasis on accuracy and predicative efficiency of the test. In the case of the ASTB, The Navy uses three different forms that measure the same outcome, this procedure is conducted for measurement in reliability and to ensure the test taker is not being administered the same exact test (NAMI, 2011). As noted earlier, the high cost for training an aviator, when the concept of utility in military aviation selection is very important and broader than validity (Cascio, 1998), it considers the costs in training, accuracy of the predictor and the importance of personnel decisions, and expenses in the selection process. As the military continues to recruit members from the different backgrounds, the notion of fairness in military aviation selection comes into sight.

E. RECENT RESEARCH

There has been a great deal of research conducted on the selection process of SNA and SNFO candidates. Many have focused on the ASTB and its prediction of attrition while others have concentrated on the reasons behind the scarcity of minorities in the aviation community. The likely explanation is the concerted effort of the Navy to implement diversity throughout every community.

> We have had great success in increasing our diversity outreach and improving diversity accessions in our ranks. We are committed to a Navy that reflects the diversity of the nation in all specialties and ranks by 2037. (Chief of Naval Operations, ADM Gary Roughead, Statement to the House Armed Forces Committee on the Department of the Navy, FY 2010)

1. Naval Aviation

A study was conducted to examine the effects of two versions of the ASTB cutoff scores on racial/ethnic minority applicants in naval aviation (Dean, 1996). This study included a data set of over 5,000 SNA applicants that entered flight training at Naval Aviation Schools Command in Pensacola, Florida from 1988 through 1994. Dean divided the data set into four groups: Caucasian, African American, Hispanic, and Asian and used flight grades from the primary phase of flight training as the determinants. A simulated cutoff score was implemented to offset the test scores for both test. The study revealed that those selected to be above the simulated cutoff score performed better, however, representation of minority groups declined. The study also showed that those performing below the simulated cutoff score experienced a higher risk of attrition.

Reinhart (1998) investigated the relationship between observable characteristics and performance in PFS. The study consisted of 276 USN Academy graduates from 1995 and 1996 that took the ASTB. It was noted that SNFOs were omitted from the study due to the difference in training curriculum. The results from the study found that the biographical inventory (BI) of the ASTB was a valid predictor for PFS completion. The study also found the PFAR, academic achievement at the Naval Academy, and previous flight experience, as valid predictors for flight training performance.

In a different study, and contrary to the Reinhart study, Wahl (1998) conducted an analysis on aviation test scores to characterize disqualification. The study sampled 2,526 SNAs and SNFOs who graduated from API and PFS from 1993 to 1997. The study examined the BI portion of the ASTB as a predictor for SNA performance in PFS. The study also looked at number of times the test was taken for success in PFS. The study found that BI was not a good predictor for flight grades and further viewed BI scores as accurate indicators for flight training disqualification. The study also concluded that SNAs who took the ASTB multiple times have a higher disqualification rate in PFS than those who passed the ASTB the first time.

Boyd (2003) conducted a study on 961 United States Naval Academy (USNA) graduates (from 1995 through 1998) to analyze determinants of student pilot success in flight training. The purpose of the study was to determine which characteristics and outcomes that are measured/determined at the Naval Academy serve as the best predictors of attrition from naval pilot training in Aviation Preflight Indoctrination (API) though Primary phase. Boyd examined the aviation assignment policy at the Naval Academy, which was composed of the ASTB and the Order of Merit (OOM), to determine if it was related to pilot performance in flight school. Alternative criteria were also looked at for developing an effective model for predicting performance. Regression models were used in conducting the analysis, including logistic regression to determine attrition rates. The results from the study showed that method used at the Naval Academy was adequate for selecting individuals for flight training and predicting attrition.

In a similar study, Gonzalez (2003) looked at predictors of the Naval Academy aviation assignment policy among graduates from 1995 through 2002. This study sampled 7,367 graduates and reviewed whether there were a difference between aviation selectees and non-aviation selectees, and pilot aviation selectees and non-pilot selectees. The results from this study showed that PFAR (an ASTB constructed score) was the most important factor in predicting aviation selection. The study also showed that PFAR and grade point average had a large impact on aviation selection and pilot selection.

Ostoin (2007) assessed the Navy's Performance-Based Measurement Battery (PBMB) that was in development and intended to supplement the ASTB. The study was conducted on 40 graduate participants with a variety of backgrounds, including 20 in aviation. The battery was administered to the participants that consisted of direction orientation tests, a dichotic listening tests, and a multi-tracking tasks. The results from the study showed that the PBMB was capable of detecting important eye-hand coordinated tracking skills and with further analysis and refinement to the scoring algorithm, this test battery should improve future aviation candidate selection.

2. Other Services

The Air Force's testing battery, the Air Force Qualifying Test, is used to select candidates for its pilot program. Researchers conducted a study to determine the effectiveness of the test and looked at the different parts of the test to see how much of an impact each would have in the selection process (Carretta, 2010). The results of the study show that these tests were accurate in determining 95% of those individuals who would excel in the flight program of which 2,190 candidates were sampled. The best predictor was those who scored well on the Academic Aptitude portion of the test, while Verbal Composite was shown to be the least effective factor in deciding who would perform well.

In another study, the U.S. Army developed a computer-based simulation test to determine what kind of deficiencies existed in their current program and how they can be corrected (Katz, 2006). The area of focus was on several different elements: perceived speed/accuracy, cognitive ability, motivation, personality and prioritization. The results were that the current testing efforts are helping to provide a foundation in determining who would be a good candidate. However, there were several changes that were recommended to the battery that includes: improving the cognitive portion of the tests, reducing the administration time, and address any kind of logistic issues that could affect testing in the future.

F. SUMMARY

After careful review of the literature, it is clear that the test battery, currently used by Naval Aviation for selecting SNAs and SNFOs, provide a basic standard for determination of which candidates have the intellect, judgment and personality necessary for success in this career field. This gives the military a foundation for testing and selection. Yet, there are a number of different problems when using this approach. The most notable include the inappropriate weighting of the questions, omission of factors that could affect the score (race/ethnicity) and analysis of performance and emotional response in combat. Consequently, it is imperative to take these different elements into consideration along with the underlying score. Once this takes place, it will provide the greatest, insights as to who would make the best aviator, and the selection process can become more efficient and refined.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODS

A. RESEARCH APPROACH

This study examined at the predictive ability of the ASTB, with respect to race and ethnicity, over the phases of the Naval Aviation Primary Flight Training. Archived test scores and training outcomes were used to determine the predictive power and effectiveness of the ASTB. Two main analyses were conducted. Multiple regressions was used to determine how effectiveness varies by racial and ethnic groups ASTB and sub-test raw scores and Navy Standard Scores (NSS), the grading and stratification standard for SNAs and SNFOs throughout flight training. This analysis identified the predictive differences between minority and majority SNA and SNFO groups' performance in Primary Flight School. Logistic regression employed ASTB sub-test raw scores and attrite status at each phase to determine the overall effectiveness of the ASTB to predict success at the end of PFS, and to predict success in Aviation Preflight Indoctrination and Introductory Flight Screening for each of the groups. This analysis identified the predictive differences between groups' completion status in each phase.

B. DATA SET

The NAMI archived data set was composed of 5,868 Naval Officers who entered service from fiscal years 2002 through 2010 through the Naval Academy, Officer Candidate School (OCS) or Naval Reserve Officer's Corps (NROTC). For the purpose of this study, given the disparate race and ethnicity sample sizes, they were split into two groupings of majority or minority for analysis. The supplied data set was received in a Microsoft Excel spreadsheet; with all personal identifying information removed. It contained 45 columns of test and demographic information, which were carefully evaluated on their level of usefulness to the study. Those with little or no bearing on the scope of the research problem were eliminated. Additionally, columns with substantial amounts of missing information were excluded (e.g., prior military service which, had data entered for fewer than half of those in the target group). Once all necessary filtering was complete, 16 factors remained to form the model: gender, race/ethnicity, design test identification, five ASTB subtest raw scores, fiscal year test taken, training pipeline, IFS attrite status, API attrite status, API NSS, PFS attrite status, PFS NSS score, and overall attrite status.

A secondary race category was displayed for some of the members in the data set supplied by NAMI. To eliminate any confusion or inexact results, a rule set was developed to determine the dominant race for those members (Table 1). Some members describe themselves as belonging to another race, and in this case, a secondary category was supplied. For the purpose of this study, a single race (dominant) was needed.

Table 1. Determining Dominant Race

Race	**Secondary Race**	**Dominant Race**
Caucasian	Hispanic	Hispanic
Caucasian	African American	African American
Hispanic	Caucasian	Hispanic
Hispanic	African American	Hispanic
African American	Hispanic	African American
African American	Caucasian	African American

SNAs and SNFOs in a not physically qualified (NPQ) status were removed from the data set. As noted, this study examined the performance of SNAs and SNFOs, members in this category did not fail from a performance measure, but from a medical or physical condition. Finally, not all data was used in this study. For example, gender was not eliminated from the data set, so that no other data points were removed when reviewing the analysis.

1. Aviation Selection Test Battery

As described earlier, the ASTB has been successfully used to screen prospective aviation candidates for the Navy, Marines Corps, and Coast Guard since World War II. All midshipmen, candidates, and civilian personnel considering a Naval Aviation career must pass this exam.

2. Navy Standard Score

The NSS, which serves as the grading and stratification standard for SNAs and SNFOs throughout flight training (up to and including PFS), will be examined as a within group and between group measure in the study and will be evaluated as a dependent variable. The Chief of Naval Aviation Training Instruction 1500.4G (2007) identifies the NSS as a representation of any score relative to the average score. The scale is artificially centered at 50 (average). Each NSS is a whole number and the scale is truncated at 20 and 80. The formula for the NSS is:

$$NSS = \left(\left(\frac{\text{grade} - \text{avg grade}}{SD}\right)10\right) + 50 \text{ rounded.}$$

Where:

grade = any student grade
avg grade = the mean grade for the distribution in question
S.D. = standard deviation for that distribution

3. Three Phases of Naval Aviation Training

Qualifying SNAs and SNFOs enter the Navy's aviation training pipeline after successful completion of the Naval Academy, OCS, or NROTC. In each phase, SNAs are rated and given a pass/fail score and, in later phases, a NSS.

Introductory Flight Screening (IFS) is the first step in the process; however, not all SNAs and SNFOs are required to complete IFS. The purpose of IFS is to screen all students to gauge their aptitude for flight in an actual aircraft, before sending them through flight school. Certain circumstances exist where an SNA did not need to attend IFS due to either: unavailability of the program, budget restraints, scheduling issues, or the student had previous flight experience. As part of the screening process at the Naval Academy and NROTC program, midshipmen are offered IFS at their schools prior to commissioning.

Aviation Preflight Indoctrination (API), also known as “ground school” is the second phase of the training program. API is a challenging six-week course that develops a foundation of aviation knowledge and skills that prepares SNAs for the demanding flight syllabus in the flying squadrons

Primary Flight School (PFS) is the third phase in flight training and teaches the SNA the basics of flying. In this phase, SNAs and SNFOs begin to fly actual military aircraft. After successful completion of PFS, SNAs and SNFOs are “Winged” as a Naval Aviator.

There is attrition among SNAs and SNFOs in every phase of the pipeline. Attrition is usually for one of three reasons:

1. Academic Failure—Academic failure includes unsatisfactory performance on classroom material and swim qualification, as this is a part of the API curriculum. Academic failure in Primary flight training is nothing more than unsatisfactory performance on a flight evolution.
2. Drop on Request (DOR)—DOR is a possibility for every SNA and SNFOs from their arrival at Naval Aviation Schools Command (NASC) until he or she is “winged” as a Naval Aviator. The most commonly reported reasons were loss of motivation and lack of desire to complete the program. This study did not involve direct contact with SNA’s and SNFOs. The subjects’ data were obtained from NAMI and all personal identifying information was removed.

3. Not Physically Qualified (NPQ)—NPQ is an indication the member has accrued some form of medical or physical condition that disqualifies him/her from the flight program.

4. Race/Ethnicity

The race/ethnicity categories included in the study are African American, Caucasian, and Hispanic. This is principally because these three groups comprise 99 percent of the SNAs and SNFOs in the data set. This correlates closely to the Department of the Navy 2010 Annual Report on Diversity.

5. Software and Hardware

Microsoft Office Excel 2007 (Smart, 2008) was used to review received data and manipulate it into a useful format. The data was then imported to JMP 9 (SAS, 2010) for producing regression models and reviewing the output information for statistical analysis. All data calculations were performed on a Dell Optiplex 380 desktop computer operating Windows 7 Professional.

C. PROCEDURES

This study was conducted in two parts. The first part examined the performance differences between minority and majority SNAs on the ASTB and the subsequent performances of these groups in PFS, as measured by the NSS. The objective of identifying a correlation between ASTB and NSS performance will allude to the predictive nature of the tool but will not prove causation.

For the second part, logistic regression analysis was performed using the additional information in the data set, from those candidates who took the ASTB, but did not complete PFS or one of the earlier phases. The second part of the study examined ASTB subtest raw scores and phase completion status. Data was analyzed to determine success in each phase and predictive power for the subgroups of African Americans, Caucasians, and Hispanics.

D. ANALYSIS

For this study, three hypotheses were examined.

1. Hypothesis One

$H1_O$: There is no difference between the predictive ability of the ASTB in minority and majority SNAs and SNFOs primary flight performance.

$H1_A$: There is a difference in the predictive nature of the ASTB for minority and majority SNAS and SNFOs flight performance.

2. Hypothesis Two

$H2_O$: There is no difference in predictive ability of the ASTB for the overall success rate at the end of PFS between minority and majority SNAs and SNFOs.

$H2_A$: There is a difference in predictive ability of the ASTB for the overall success rate at the end of PFS between minority and majority SNAs and SNFOs.

3. Hypothesis Three

$H3_O$: There is no difference in predictive ability of the ASTB for success in the earlier phases of flight training (API and IFS) between minority and majority SNAs and SNFOs.

$H3_A$: There is a difference in predictive ability of the ASTB for success in the earlier phases of flight training (API and IFS) for between and majority SNAs and SNFOs.

To test the first hypothesis, stepwise multiple regressions were run, adding and/or removing each dependent variable in the interest of making a determination of which independent variables (IVs) are the best predictors of performance in flight school regression models of the form:

$$y_t = \beta_0 + \beta_1 \cdot x_{1t} + \beta_2 \cdot x_{2t} + \cdots + \beta_n \cdot x_{nt} + \epsilon_t$$

were generated to determine the best set of predictors for the minority and majority groups. Each model was then fit to the entire population to determine if a significant difference exists between the overall population model and the group models.

For the regression analysis, the subtests of the ASTB (MST, RCT, MCT, SAT, and ANIT) served as independent variables. NSS was used as the dependent variable, as noted earlier it is a quantitative expression of performance in PFS.

The second and third hypotheses examined the predicative ability of the ASTB in the three phases of the flight training program: IFS, API, and PFS. Logistic regression (also known as logistic or logit model) was used to determine which of the subtests are the best predictors for success in each phase in the flight training program. Logistic regression is a powerful extension of multiple regression when the dependent variable is categorical (either 0 or 1) (Norman & Streiner, 2003). It works by computing a logistic function (Figure 1.) from the predictor variables and then comparing the computed probabilities to 1s and 0s.

Logistic function formula:

$$f(y) = \frac{1}{1 + e^{-y}}$$

where the value of y is derived from the multiple regression model. An example of this model, as depicted in figure 1, would predict if somebody with an ASTB score of 2 to have an 85% probability of success.

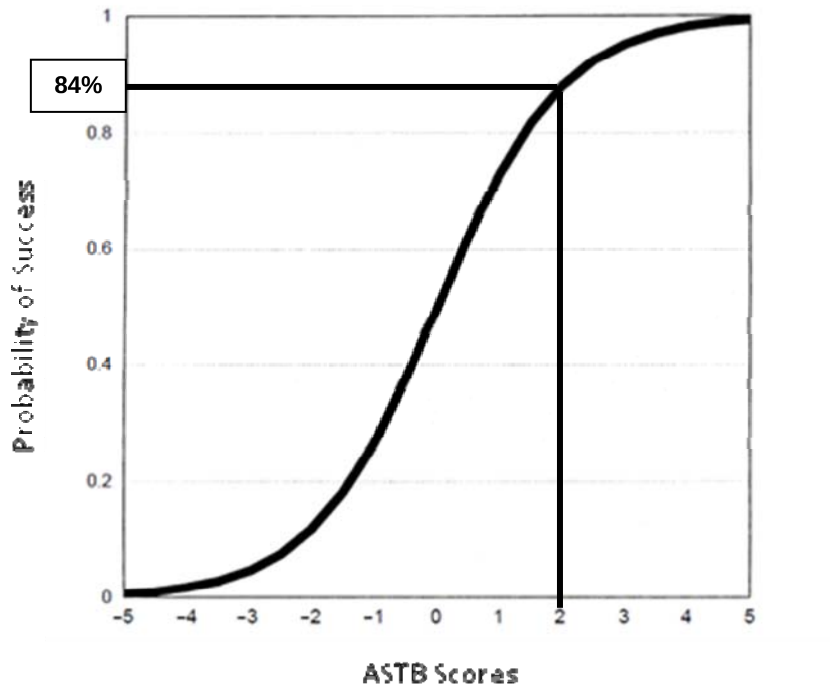


Figure 1. Logistic Function Graph

Models were generated to determine the best set of predictors for each of the groups. Each model was then compared between the groups and with the overall model to determine the difference and predictive power for success.

For the logistic regression analysis, the subtests of the ASTB (MST, RCT, MCT, SAT, and ANIT) served as the independent variables. Attrite status of the three phases (IFS, API and PFS) was used as the dependent variable, which is indicated for this study by either a 0 for failure to complete the phase or a 1 for successful completing the phase.

IV. RESULTS

The data set 5,868 SNAs and SNFOs candidates was filtered and subsequently split into majority (n= 2,910) and minority (n= 329) groupings for analysis. The minority group was further broken down into African American (n= 91) and Hispanic (n= 238) subgroups for subsequent analysis. Data was organized using Microsoft Excel (Smart, 2008), and analyzed using JMP 9.

A. MULTIPLE REGRESSION

To test the first hypothesis, a linear regression was run on the NSS dependent variable—NSS. The objective was to discover which, if any, of the five ASTB subtests have any predictive power for the students' performance in primary flight training. Linear regression analyses were conducted in JMP 9, for the entire sample as well as each of the racial/ethnic subgroups.

1. Minority Group Results

a. African Americans

The amount of data available was limited by the fact that there are not very many African Americans in the Naval Aviation relative to the total number of SNAs (n=17, 3%). No ASTB subtests were significant ($p > 0.05$) with respect to predicting the success (i.e., high NSS) of the African American students (see Figure 2 and Table 2). Full model results are available in Appendix A.

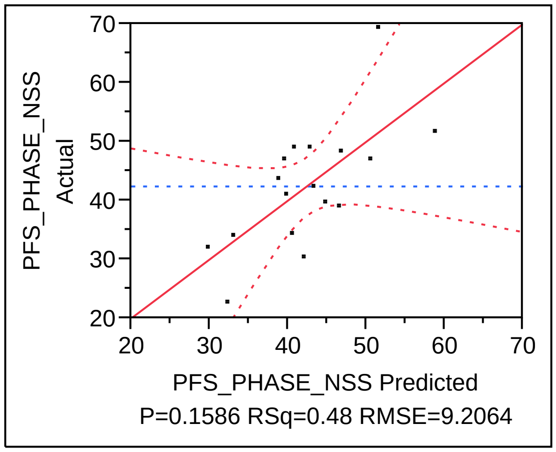


Figure 2. Response PFS_PHASE_NSS Race=Afri

Table 2. Multiple Regression for African American SNAs

RSquare	0.475019
RSquare Adj	0.236391
Root Mean Square Error	9.206366
Mean of Response	42.40588
Observations (or Sum Wgts)	17

Parameter estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	34.382964	4.328523	7.94	<.0001*
ANI_RAW	8.5030891	4.465524	1.90	0.0834
MCT_RAW	7.7869447	5.469825	1.42	0.1823
MST_RAW	-0.56164	4.668826	-0.12	0.9064
RCT_RAW	-1.265398	5.068415	-0.25	0.8074
SAT_RAW	2.9931577	3.283848	0.91	0.3816

b. Hispanics

Similar to the problem that was encountered with the African American sample, the data set did not contain enough observations from those in the Hispanic group data to formulate a model with any predictive power. The Hispanic population in the aviation training pipeline during the time that this data was gathered was very small (n=19, 5%). In the analysis, none of the ASTB subtests were found to be significant for generating a prediction model for

Hispanics' primary flight performance, as measured by NSS (see Figure 3 and Table 3). Full model results are available in Appendix A.

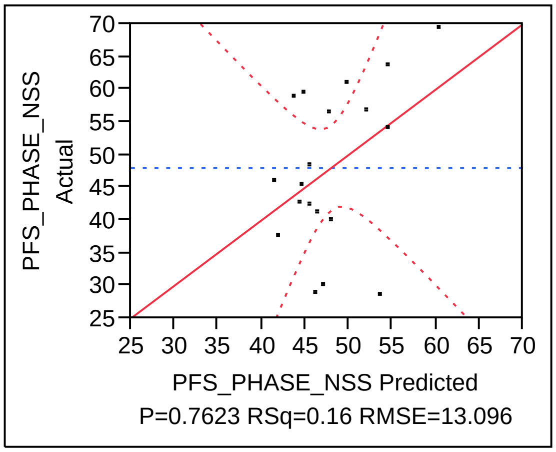


Figure 3. Response PFS_PHASE_NSS Race=Hisp

Table 3. Multiple Regression for Hispanic SNAs

RSquare	0.164644
RSquare Adj	-0.15665
Root Mean Square Error	13.09568
Mean of Response	47.97895
Observations (or Sum Wgts)	19

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	42.707525	5.226619	8.17	<.0001*
ANI_RAW	0.5912756	5.969145	0.10	0.9226
MCT_RAW	-2.320762	8.69574	-0.27	0.7937
MST_RAW	5.0984412	6.049959	0.84	0.4146
RCT_RAW	2.6853653	7.071583	0.38	0.7103
SAT_RAW	5.4299217	7.130048	0.76	0.4599

2. Majority Group Results

Accounting for approximately 93% of the data available for this analysis, the majority sample (n=493) yielded much cleaner results. The analysis revealed that the ANI, MCT, RCT and SAT were all significant predictors of the majority students' NSS scores at the conclusion of Primary flight training (see Figure 4 Table 4). Furthermore, all of the subtests for this group displayed directly proportional (i.e. positive) qualities with respect to the NSS. The R^2 (0.1241) and the Adjusted R^2 (0.1151), suggest that the model does not account for a great deal of the variability in the sample, but the low probabilities produced in the analysis (Prob > |t| 0.0061 ANI, 0.0258 MCT, 0.0032 RCT, and 0.0017 SAT) validate the fact that significance of these tests in predicting performance is not random, even if a very small level of significance is chosen. Full model results are available in Appendix A.

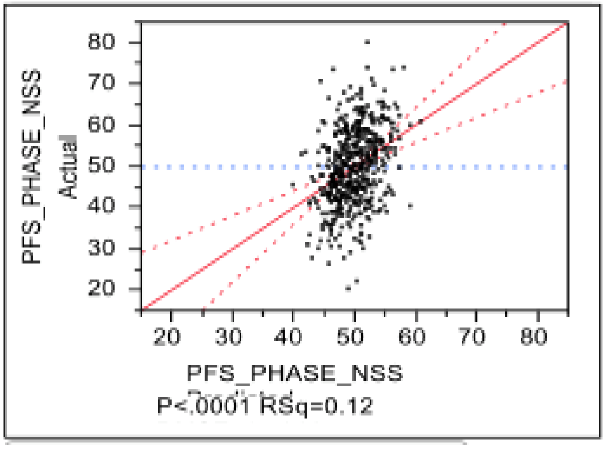


Figure 4. Response PFS_PHASE_NSS Race=Caucasian

Table 4. Regression for Majority SNAs

RSquare	0.124103
RSquare Adj	0.115111
Root Mean Square Error	8.786242
Mean of Response	49.86471
Observations (or Sum Wgts)	493

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	44.355137	0.788347	56.26	<.0001*
ANI_RAW	2.2385838	0.811977	2.76	0.0061 *
MCT_RAW	1.6109113	0.720243	2.24	0.0258 *
MST_RAW	1.2602277	0.702181	1.79	0.0733
RCT_RAW	2.3325161	0.787373	2.96	0.0032 *
SAT_RAW	1.980858	0.628745	3.15	0.0017 *

B. LOGISTIC REGRESSION

For the second hypothesis, an overall model of the groups was developed for each of the three phases using logistic regression. After an overall model was developed, logistic regression was conducted again for each of the race/ethnicity groupings; the results are shown for each of the phases in the following sections.

1. PFS Results

As mentioned earlier, PFS is the third phase in the Naval Aviation Flight Training Program; members enter this phase after successfully completing API. For this portion of the study, there were 739 observations, of which 21 were African Americans (3%), 680 were Caucasians (92%) and 38 were Hispanics (5%). Even though PFS is the third phase in the order of the training pipeline, it was contended that these results should be explained first since this population is closest to the one considered in the earlier multiple regression analysis.

For the overall model, there was nothing significant to report. None of the IVs were found to be statistically significant as a predictor of success in PFS (see Table 5). All p-values (under the Prob/ChiSq column) are significantly larger than the chosen level of significance of 0.05. In addition, the R Square value of 0.0213 indicates that this model only explains 2% of the variation in observed success in PFS.

Table 5. Overall PFS Logistic Regression Model Results

Overall Group PFS			**Prob/ChiSq**
R Square	0.0213		0.2047
Observations	739		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.2319	<.0001*
ANIT_RAW		-0.2438	0.4132
MCT_RAW		0.2319	0.4253
MST_RAW		-0.2236	0.4120
RCT_RAW		-0.3373	0.2653
SAT_RAW		-0.4392	0.0634

For the subgroups, a model could not be produced for African Americans because all members from this group who entered PFS passed. Models were produced for Caucasians and Hispanics, but none of the IVs were found to be statistically significant as a predictor of success in PFS (see Table 6). The model for Caucasians was found to be similar to the overall group model, but this was due to the fact that this group makes up 92% of the overall group. Additional output is located in Appendix B.

Table 6. Group PFS Logistic Regression Model Results

Caucasian			**Prob/ChiSq**
R Square	0.1378		0.2425
Observations	680		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.1653	<.0001*
ANIT_RAW		-0.2267	0.4590
MCT_RAW		0.1566	0.5965
MST_RAW		-0.2165	0.4343
RCT_RAW		-0.3164	0.3066
SAT_RAW		-0.4373	0.0719

Hispanic			**Prob/ChiSq**
R Square	0.1591		0.7774
Observations	38		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.8198	0.0113*
ANIT_RAW		-0.1560	0.9318
MCT_RAW		2.2133	0.3484
MST_RAW		-1.7520	0.4110
RCT_RAW		-0.9695	0.5958
SAT_RAW		-1.5948	0.3483

2. API Results

The second phase in the training pipeline is API. Also known as "ground School", this is where the majority of SNAs begin their flying careers. For API, there were 3,093 observations, of which 84 were African Americans (3%), 2,795 were Caucasians (90%) and 214 were Hispanics (7%).

The overall group output model found that ANIT, MCT, MST, and SAT were statistically significant IVs (see Table 7). The p-values for these variables were lower than the level of significance of 0.05. The R Square value of 0.1393 indicates that this model only explains 14% of the variation in observed success in API. Logistic regression produced the following model:

Y = -2.4810 -1.1118 ANIT -0.7025 MCT -0.8508 MST -0.6282 SAT

The coefficients for this model come out to be negative from the results of logistic regression. This is nothing to be concerned about; the logistic function transformation described in the previous chapter will make the probability a positive number between "0" and "1."

Table 7. Overall API Logistic Regression Model Results

Overall Group API			**Prob/ChiSq**
R Square	0.1378		<.0001*
Observations	3093		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.4810	<.0001*
ANIT_RAW		-1.1118	<.0001*
MCT_RAW		-0.7025	0.0016*
MST_RAW		-0.8508	0.0001*
RCT_RAW		-0.3345	0.1561
SAT_RAW		-0.6282	0.0010*

In API, logistic regression results found subtests to be statistically significant for Caucasians and Hispanics (see Table 8). Of interest, results for each group produced a model that was different from one another. There was nothing significant to report for African Americans. However, R Square for this subgroup was higher than the overall model with 18% variation explained by the model, but this was due to the small sample size (84 observations) of the group. The model for Caucasians resembled the overall model because this group makes up 90% of the total samples observed. For Hispanics, MST was the only IV that was significant. This model was different from the overall group and Caucasians. Hispanics made up 7% of the observations but produced an R Square of 15%, which is slightly higher than the overall model. Additional output is located in Appendix C.

Table 8. Group API Logistic Regression Model Results

African Americans			**Prob/ChiSq**
R Square	0.1776		0.0709
Observations	84		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-1.6372	0.0025*
ANIT_RAW		-0.4900	0.5399
MCT_RAW		-0.2522	0.7399
MST_RAW		-1.2274	0.1508
RCT_RAW		0.1740	0.8281
SAT_RAW		-1.2976	0.0927

Caucasians			**Prob/ChiSq**
R Square	0.1319		<.001*
Observations	2795		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.4810	<.0001*
ANIT_RAW		-1.1118	<.0001*
MCT_RAW		-0.7025	0.0144*
MST_RAW		-0.8508	0.0120*
RCT_RAW		-0.3345	0.1190
SAT_RAW		-0.6282	0.0031*

Hispanics			**Prob/ChiSq**
R Square	0.1524		<.0001*
Observations	214		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-3.1436	<.0001*
ANIT_RAW		-0.6984	0.3432
MCT_RAW		-1.1606	0.0979
MST_RAW		-1.6818	0.0243*
RCT_RAW		0.5184	0.4132
SAT_RAW		0.1395	0.8042

The graph of the API logistic regression results for the MST subtest of the ASTB shows the results for each group and the overall population. Each curve predicts the probability of success for that group. An example of this would be if a person were to score a negative one (-1) on the MST, what would be the probability of success if he/she belonged to one of the groups. The results from this would include:

African American = (1-.33) = 67% percent of success.

Caucasian = 93% percent of success.

Hispanic = 88% percent of success.

Overall = 91% percent of success.

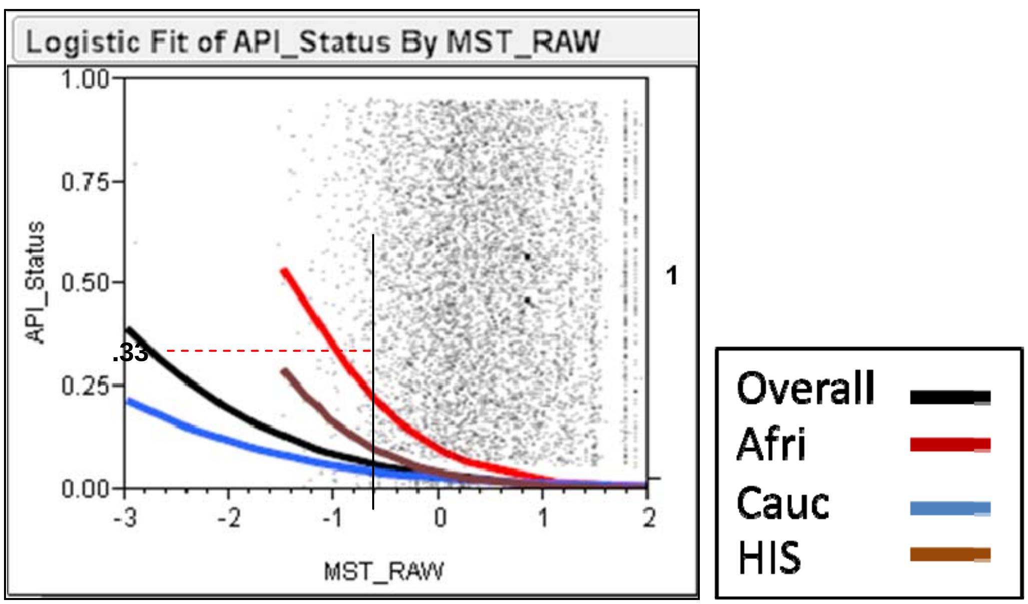


Figure 5. API Logistic Results for MST

3. IFS Results

The first phase in the training pipeline is IFS, and as noted earlier, not all SNAs and SNFOs are required to attend IFS, but Naval Academy and ROTC midshipmen are offered IFS prior to commissioning. For IFS, there were 3,239 observations, of which 91 were African Americans (3%), 2,910 were Caucasians (90%), and 238 were Hispanics (7%).

The overall group model found that there were statistically significant IVs for MCT and SAT (see Table 9). The p-values for these variables were lower than the level of significance of 0.05. R- Square for the overall model was lowest of the phases with only 2% of the variance explained by the model. Logistic regression produced the following model:

Y = - 2.6873 - 0.3214 MCT – 0.2680 SAT

Table 9. Overall IFS Logistic Regression Model Results

Overall Group IFS			**Prob/ChiSq**
R Square	0.0161		0.0023*
Observations	3239		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.6873	<.0001*
ANIT_RAW		-0.1419	0.4475
MCT_RAW		-0.3215	0.0368*
MST_RAW		-0.1778	0.2348
RCT_RAW		-0.0162	0.9251
SAT_RAW		-0.2680	0.0470*

The results for IFS were similar to API; logistic regression produced statistically significant results for only two of the groups (see Table 10). There was nothing significant to report for African Americans. Caucasians made up 90% of the total samples observed and produced a model similar to the overall model. For Hispanics, MST was again the only IV that was significant. Additional output is located in Appendix D.

Table 10. Group IFS Logistic Regression Model Results

African Americans			**Prob/ChiSq**
R Square	0.0467		0.9091
Observations	91		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-1.6372	<.0001*
ANIT_RAW		-0.4900	0.6647
MCT_RAW		-0.2522	0.6789
MST_RAW		-1.2274	0.7120
RCT_RAW		0.1740	0.4516
SAT_RAW		-1.2976	0.7192

Caucasians			**Prob/ChiSq**
R Square	0.0208		0.0007*
Observations	2910		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-2.4810	<.0001*
ANIT_RAW		-1.1118	0.6202
MCT_RAW		-0.7025	0.0085*
MST_RAW		-0.8508	0.4861
RCT_RAW		-0.3345	0.6973
SAT_RAW		-0.6282	0.0294*

Hispanics			**Prob/ChiSq**
R Square	0.0670		0.3093
Observations	238		
Variable		**Estimate**	**Prob/ChiSq**
Intercept		-3.1436	<.0001*
ANIT_RAW		-0.6984	0.5987
MCT_RAW		-1.1606	0.1234
MST_RAW		-1.6818	0.0398*
RCT_RAW		0.5184	0.8843
SAT_RAW		0.1395	0.9637

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSION AND RECOMMENDATIONS

This study looked at the predictive ability of the ASTB with respect to race/ethnicity throughout the different phases of the Naval Aviation Primary Flight Training. Two analyses were conducted: 1) to determine if there was a difference between majority and minority group performance in PFS; and 2) to determine how well the ASTB could predict success in each training phase for three groups—African Americans, Caucasians, and Hispanics.

A. MULTIPLE REGRESSION CONCLUSIONS

Multiple regressions was used to determine if a difference exists between the majority and minority groups with respect to the predictive ability of the ASTB subtests. The subtest scores served as the independent variables while the students' respective NSSs served as the dependent variable.

1. African American

Due to the extremely small size of this segment of the data set, we were unable to fit a predictive model to the African American group. None of the subtests yielded any amount of statistical significance.

2. Caucasian

The data for the Caucasian sample represented the vast majority of the SNFOs and SNAs who matriculated through the pipeline during FY 2002–FY 2010. A model was successfully fit to the majority group in which all subtests were statistically significant with the exception of the MST. Once the model was created, it was fit to the entire dataset with the subtests and race as independent variables. This produced a model that displayed statistical significance for every independent variable, including race.

3. Hispanic

The Hispanic sample presented challenges similar those observed during the analysis of the African American group. None of the subtests were significant, and no predictive model was derived from the data provided.

4. Hypothesis One Conclusion

The null for Hypothesis One was not rejected and it is concluded that there is no difference between the predictive ability of the ASTB in minority and majority SNAs primary flight performance. This determination was made because it could not be reasonably concluded that there is a significant difference between the NSS for the majority and minority groups without a model to substantiate that assertion.

B. LOGISTIC REGRESSION CONCLUSIONS

This analysis examined the ASTB subtest raw scores as the independent variables and attrite status as the dependent variable.

1. African Americans

For African Americans, it was difficult to form any conclusions due to the small sample sizes in each phase, African Americans only made up 3% of the observations in all three phases of the training pipeline. In PFS, no model could be produced due to the 100% success rate when they entered this phase. The results showed that there was nothing significant to report for API and IFS. None of the IVs from the ASTB were good predictors of success for this group.

2. Caucasians

For each of the phases, Caucasians made up the majority of the groups with 92% in PFS, and 90% in both API and IFS. Due to the high percentage of the overall sample in each phase, the models for Caucasians resembled the overall model. In PFS, the results from the analysis showed that there was nothing significant to report for this group and the IVs were not good predictors for success in this phase. For API, the logistic regression results showed that the variables from the ASTB were better predictors here than the other two phases.

3. Hispanics

Hispanics made up 5% in PFS, and 7% in API and IFS. Logistic regression produced models in API and IFS that were different from the overall model and the other groups; however, there was only one significant predictor for each of the models.

4. Hypothesis Two Conclusion

There is no evidence to conclude that we should reject the null hypothesis. The results from the logistic regression models found that there was nothing significant to report from the overall and the subgroup models. The results from the logistic regression models do not show that the ASTB is a good predictor of success for PFS.

5. Hypothesis Three Conclusion

In API, there was evidence that four of the subtests were significant and positive predictors in the overall model and for Caucasians; however, these models only explained a small proportion of the total variation, with an R Square of 14% and 13%, respectively. The model for Hispanics only showed one of the four predictors from the overall model as being significant; however, the R Square of 15% was only slightly larger than the overall model.

The results for IFS showed that there were two positive predictors in the overall model and for Caucasians, and only one predictor for Hispanics. However, the R Square for all three were very low with 2% of the variation explained for the overall model and Caucasians, and 7% for the Hispanics.

There is evidence to conclude that we should reject the null hypothesis, however, the predictive power of the overall model was small for API and IFS, and the predictive power for the subgroups was no better than the overall model. We conclude that the ASTB is not a good predictor for the earlier phases of flight training.

C. RECOMMENDATIONS

An additional linear regression analysis should be conducted when the minority representation in the data set is larger. Additionally, there were numerous observations from students that finished the pipeline, but had no NSS entered into the spreadsheet and were subsequently unusable in this analysis. Efforts to collect that data should be made by NAMI; it may be available in individual personnel records.

Logistic Regression was only used to study the predictive ability and success for three groups in PFS, and the earlier phases (API and IFS) of flight training. It was apparent that the small data set for minorities limited our ability in the findings. Of the minority groups, only Hispanics were able to produce a model different from the rest. This group can be further investigated to see how and why these findings were different from the other groups and the overall model.

It is also recommended that research be conducted with the retained data and include all minority groups, including separating males and females (even though females make up a very small percentage) to compare and examine how much predicative ability the ASTB would have on all potential SNAs and SNFOs.

As additional data becomes available, further research can also be conducted to determine predictive power and success for fiscal year groups, test versions and SNAs age at time of test.

APPENDIX A. LINEAR REGRESSION MODEL RESULTS

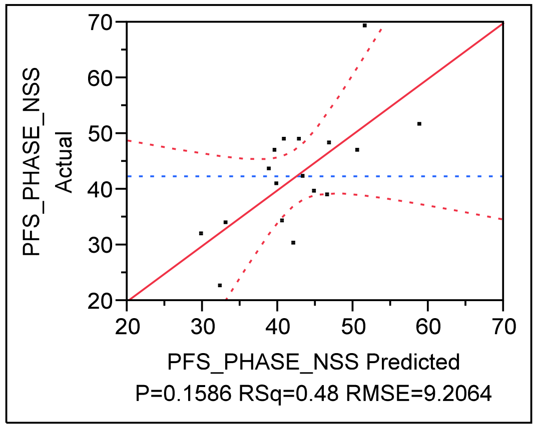


Figure 6. Response PFS_PHASE_NSS Race=African American

Table 11. Response PFS_PHASE_NSS Race=African American

RSquare	0.475019
RSquare Adj	0.236391
Root Mean Square Error	9.206366
Mean of Response	42.40588
Observations (or Sum Wgts)	17

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	843.6004	168.720	1.9906
Error	11	932.3290	84.757	**Prob > F**
C. Total	16	1775.9294		0.1586

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.382964	4.328523	7.94	<.0001*
ANI_RAW	8.5030891	4.465524	1.90	0.0834
MCT_RAW	7.7869447	5.469825	1.42	0.1823
MST_RAW	-0.56164	4.668826	-0.12	0.9064
RCT_RAW	-1.265398	5.068415	-0.25	0.8074
SAT_RAW	2.9931577	3.283848	0.91	0.3816

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	t Ratio	Prob>|t|
ANI_RAW	8.5030891	4.465524	1.90		0.0834
MCT_RAW	7.7869447	5.469825	1.42		0.1823
SAT_RAW	2.9931577	3.283848	0.91		0.3816
RCT_RAW	-1.265398	5.068415	-0.25		0.8074
MST_RAW	-0.56164	4.668826	-0.12		0.9064

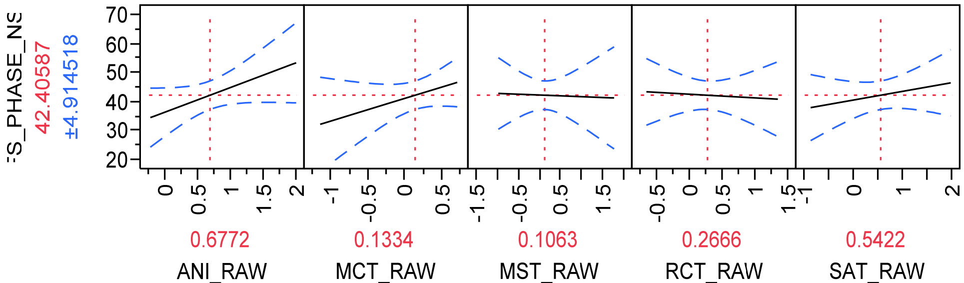


Figure 7. Prediction Profiler for African American

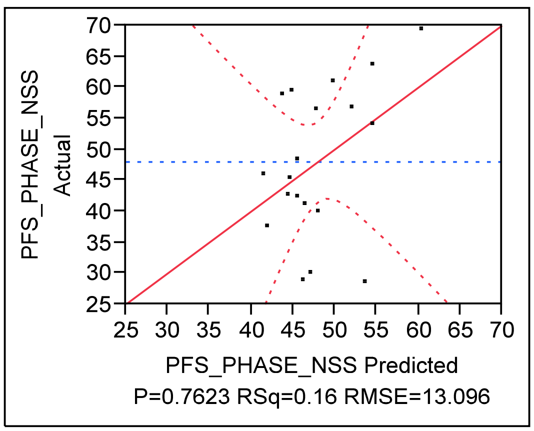


Figure 8. Response PFS_PHASE_NSS Race=Hispanic

Table 12. Response PFS_PHASE_NSS Race=Hispanic

RSquare	0.164644
RSquare Adj	-0.15665
Root Mean Square Error	13.09568
Mean of Response	47.97895
Observations (or Sum Wgts)	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	439.4125	87.883	0.5124
Error	13	2229.4590	171.497	**Prob > F**
C. Total	18	2668.8716		0.7623

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	42.707525	5.226619	8.17	<.0001*
ANI_RAW	0.5912756	5.969145	0.10	0.9226
MCT_RAW	-2.320762	8.69574	-0.27	0.7937
MST_RAW	5.0984412	6.049959	0.84	0.4146
RCT_RAW	2.6853653	7.071583	0.38	0.7103
SAT_RAW	5.4299217	7.130048	0.76	0.4599

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	t Ratio	Prob>|t|
MST_RAW	5.0984412	6.049959	0.84		0.4146
SAT_RAW	5.4299217	7.130048	0.76		0.4599
RCT_RAW	2.6853653	7.071583	0.38		0.7103
MCT_RAW	-2.320762	8.69574	-0.27		0.7937
ANI_RAW	0.5912756	5.969145	0.10		0.9226

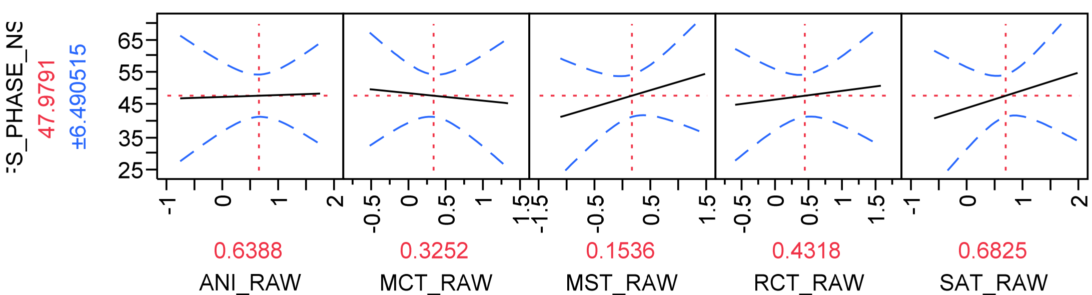
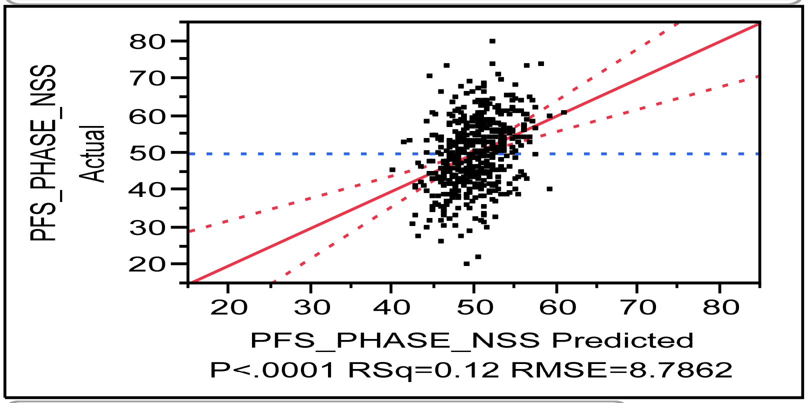


Figure 9. Prediction Profiler for Hispanic

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.124103
RSquare Adj	0.115111
Root Mean Square Error	8.786242
Mean of Response	49.86471
Observations (or Sum Wgts)	493

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	5326.794	1065.36	13.8003
Error	487	37595.452	77.20	**Prob > F**
C. Total	492	42922.246		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	44.355137	0.788347	56.26	<.0001*
ANI_RAW	2.2385838	0.811977	2.76	0.0061*
MCT_RAW	1.6109113	0.720243	2.24	0.0258*
MST_RAW	1.2602277	0.702181	1.79	0.0733
RCT_RAW	2.3325161	0.787373	2.96	0.0032*
SAT_RAW	1.980858	0.628745	3.15	0.0017*

Residual by Predicted Plot

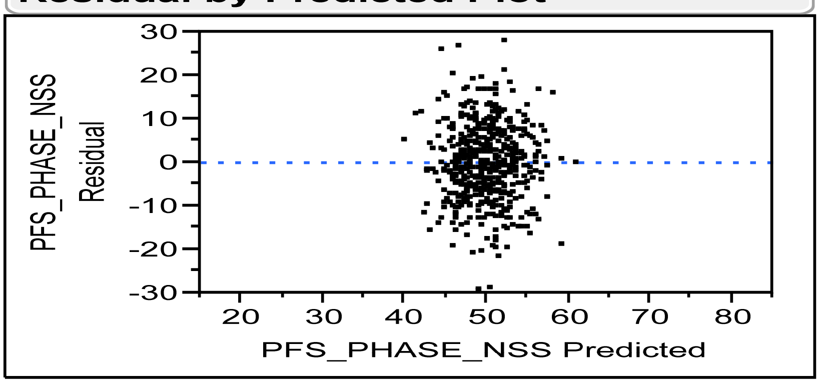
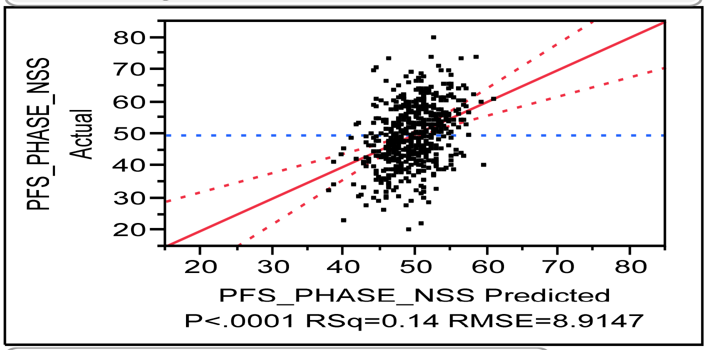


Figure 10. Response PFS_PHASE_NSS Race=Caucasian

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.143301
RSquare Adj	0.131791
Root Mean Square Error	8.914672
Mean of Response	49.55728
Observations (or Sum Wgts)	529

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	6925.806	989.401	12.4498
Error	521	41404.588	79.471	**Prob > F**
C. Total	528	48330.394		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	42.077806	1.171914	35.91	<.0001*
ANI_RAW	2.5731579	0.785784	3.27	0.0011*
MCT_RAW	1.6135399	0.712235	2.27	0.0239*
MST_RAW	1.3794799	0.684419	2.02	0.0444*
RCT_RAW	2.1263073	0.768368	2.77	0.0059*
SAT_RAW	1.9686688	0.613144	3.21	0.0014*
Race[Afri]	-3.410632	1.609694	-2.12	0.0346*
Race[Cauc]	2.1515286	1.046728	2.06	0.0403*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
ANI_RAW	1	1	852.19242	10.7233	0.0011*
MCT_RAW	1	1	407.87126	5.1323	0.0239*
MST_RAW	1	1	322.84740	4.0624	0.0444*
RCT_RAW	1	1	608.58969	7.6580	0.0059*
SAT_RAW	1	1	819.27788	10.3091	0.0014*
Race	2	2	501.49646	3.1552	0.0434*

Residual by Predicted Plot

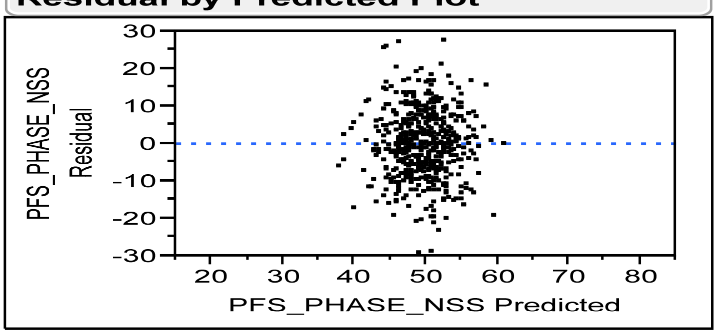


Figure 11. Response PFS_PHASE_NSS ALL STUDENTS

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. PFS LOGISTIC REGRESSION MODEL RESULTS AND GRAPHS

Table 13. Overall PFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	3.61051	5	7.221019	0.2047
Full	165.92930			
Reduced	169.53981			

RSquare (U)	0.0213
Observations (or Sum Wgts)	739

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.2319386	0.2527363	77.99	<.0001*
ANIT_RAW	-0.2438056	0.2979695	0.67	0.4132
MCT_RAW	0.23192432	0.2909327	0.64	0.4253
MST_RAW	-0.2236322	0.2726034	0.67	0.4120
RCT_RAW	-0.3372685	0.302792	1.24	0.2653
SAT_RAW	-0.4391577	0.2365685	3.45	0.0634

Table 14. Caucasian PFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	3.35893	5	6.717858	0.2425
Full	156.97030			
Reduced	160.32923			

RSquare (U)	0.0210
Observations (or Sum Wgts)	680

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.1652756	0.2635056	67.52	<.0001*
ANIT_RAW	-0.2267403	0.306207	0.55	0.4590
MCT_RAW	0.15664854	0.2958406	0.28	0.5965
MST_RAW	-0.2165378	0.2769443	0.61	0.4343
RCT_RAW	-0.3164083	0.3094831	1.05	0.3066
SAT_RAW	-0.4373261	0.2429992	3.24	0.0719

Table 15. Hispanic PFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1.2468550	5	2.49371	0.7774
Full	6.5884429			
Reduced	7.8352979			

RSquare (U)	0.1591
Observations (or Sum Wgts)	38

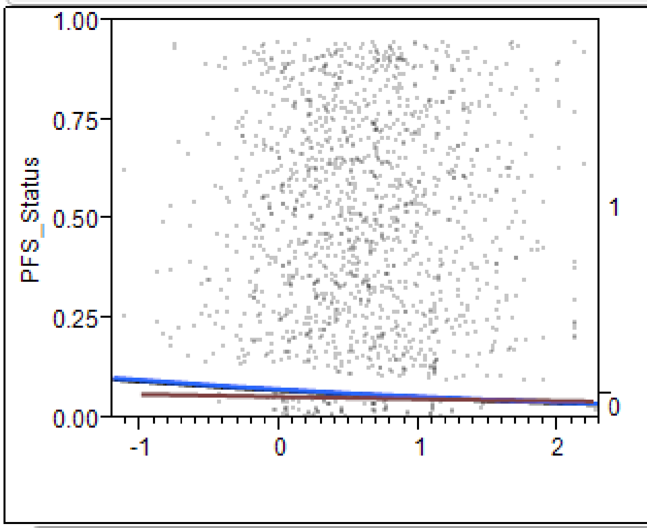
Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.8198121	1.1130826	6.42	0.0113*
ANIT_RAW	-0.1559712	1.8231615	0.01	0.9318
MCT_RAW	2.21331104	2.360594	0.88	0.3484
MST_RAW	-1.7519849	2.13097	0.68	0.4110
RCT_RAW	-0.9694933	1.8276849	0.28	0.5958
SAT_RAW	-1.5948206	1.7003372	0.88	0.3483

Fit Y by X Group

Logistic Fit of PFS_Status By ANIT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.60983	1	1.219669	0.2694
Full	168.92998			
Reduced	169.53981			

RSquare (U)	0.0036
Observations (or Sum Wgts)	739

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.567574	0.2092685	150.54	<.0001*
ANIT_RAW	-0.3041858	0.2771755	1.20	0.2724

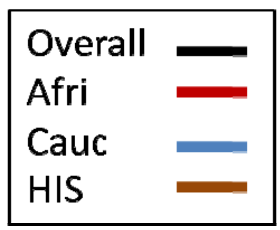
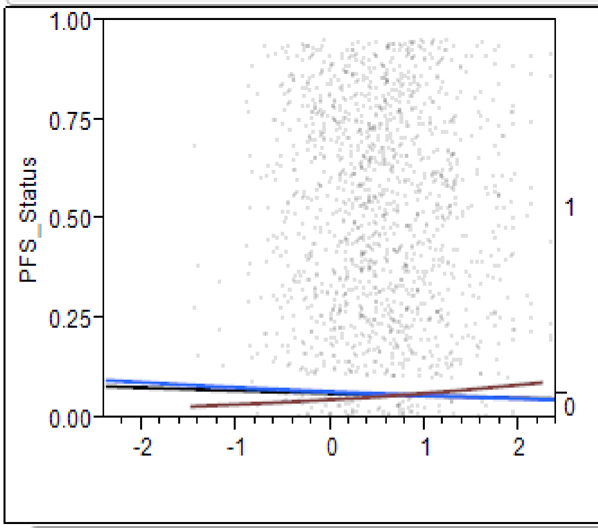


Figure 12. Group PFS Model for ANIT

Logistic Fit of PFS_Status By MCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.10058	1	0.201163	0.6538
Full	169.43923			
Reduced	169.53981			

RSquare (U)	0.0006
Observations (or Sum Wgts)	739

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.6825863	0.1916014	196.02	<.0001*
MCT_RAW	-0.1085055	0.2423391	0.20	0.6543

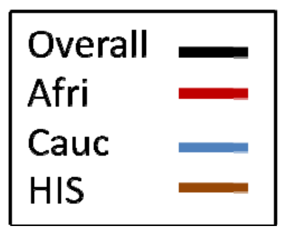
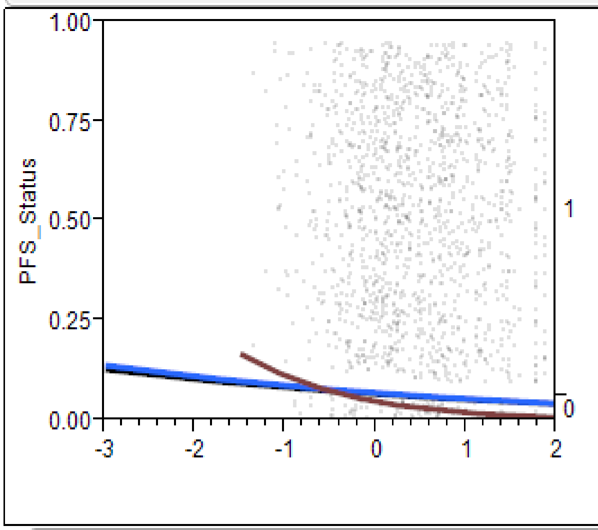


Figure 13. Group PFS Model for MCT

Logistic Fit of PFS_Status By MST_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.53552	1	1.071043	0.3007
Full	169.00429			
Reduced	169.53981			

RSquare (U)	0.0032
Observations (or Sum Wgts)	739

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.6451643	0.1722114	235.93	<.0001*
MST_RAW	-0.2426156	0.2355493	1.06	0.3030

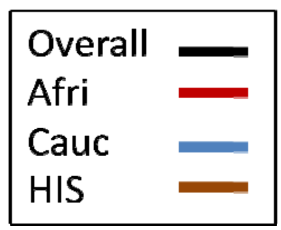
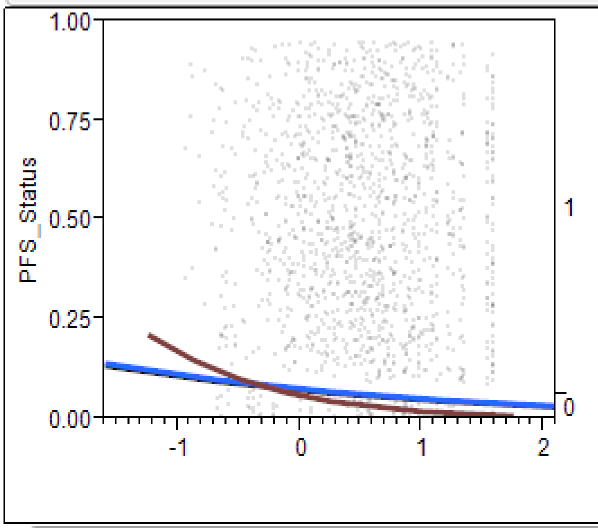


Figure 14. Group PFS Model for MST

Logistic Fit of PFS_Status By RCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1.18593	1	2.371864	0.1235
Full	168.35388			
Reduced	169.53981			

RSquare (U)	0.0070
Observations (or Sum Wgts)	739

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.5459735	0.1878649	183.66	<.0001*
RCT_RAW	-0.4278265	0.2781746	2.37	0.1241

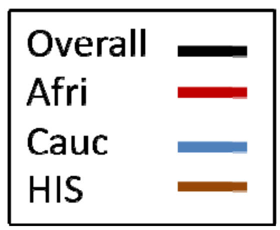
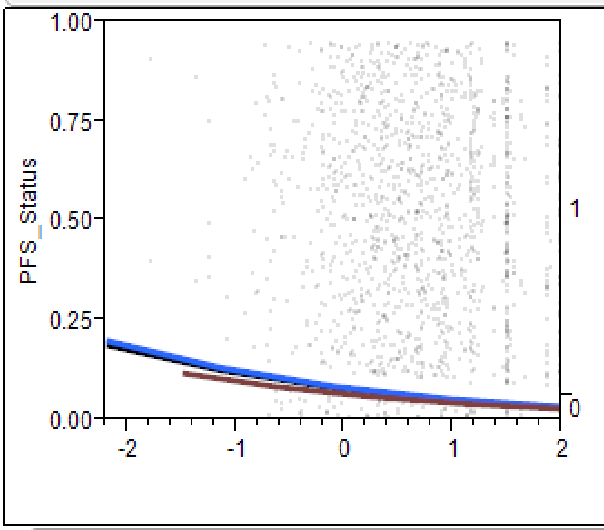


Figure 15. Group PFS Model for RCT

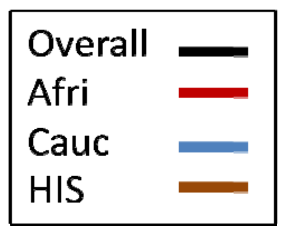
Logistic Fit of PFS_Status By SAT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2.04279	1	4.085576	0.0433*
Full	167.49702			
Reduced	169.53981			

RSquare (U)	0.0120
Observations (or Sum Wgts)	739



Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.4730438	0.1890658	171.10	<.0001*
SAT_RAW	-0.4609948	0.2282072	4.08	0.0434*

Figure 16. Group PFS Model for SAT

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. API LOGISTIC REGRESSION MODEL RESULTS AND GRAPHS

Table 16. Overall API Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	49.14969	5	98.29938	<.0001*
Full	307.57727			
Reduced	356.72696			

RSquare (U)	0.1378
Observations (or Sum Wgts)	3093

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.4809712	0.1504755	271.84	<.0001*
ANIT_RAW	-1.1117785	0.2581364	18.55	<.0001*
MCT_RAW	-0.7025331	0.2225352	9.97	0.0016*
MST_RAW	-0.8508375	0.2198958	14.97	0.0001*
RCT_RAW	-0.3345054	0.2358754	2.01	0.1561
SAT_RAW	-0.6282367	0.1909254	10.83	0.0010*

Table 17. African American API Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	5.079021	5	10.15804	0.0709
Full	23.522961			
Reduced	28.601981			

RSquare (U)	0.1776
Observations (or Sum Wgts)	84

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-1.6371661	0.5405978	9.17	0.0025*
ANIT_RAW	-0.4900352	0.7994206	0.38	0.5399
MCT_RAW	-0.2522312	0.7596356	0.11	0.7399
MST_RAW	-1.2274054	0.8543596	2.06	0.1508
RCT_RAW	0.17400822	0.8012856	0.05	0.8281
SAT_RAW	-1.2976476	0.7717512	2.83	0.0927

Table 18. Caucasian API Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	36.70570	5	73.4114	<.0001*
Full	241.58365			
Reduced	278.28935			

RSquare (U)	0.1319
Observations (or Sum Wgts)	2795

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.5527027	0.1716976	221.04	<.0001*
ANIT_RAW	-1.3062389	0.3006137	18.88	<.0001*
MCT_RAW	-0.6211053	0.2538787	5.99	0.0144*
MST_RAW	-0.6250011	0.2486944	6.32	0.0120*
RCT_RAW	-0.4258228	0.2731629	2.43	0.1190
SAT_RAW	-0.6434701	0.2177656	8.73	0.0031*

Table 19. Hispanic API Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.157343	5	12.31469	0.0307*
Full	34.239194			
Reduced	40.396538			

RSquare (U)	0.1524
Observations (or Sum Wgts)	214

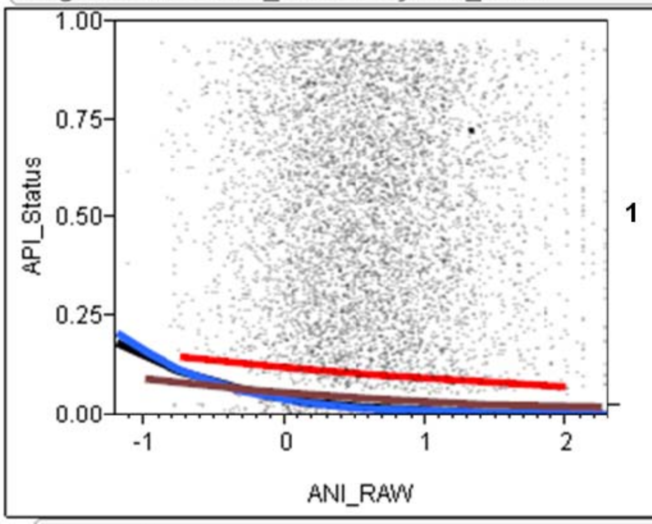
Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.1435644	0.5764101	29.74	<.0001*
ANI_RAW	-0.6984275	0.7368486	0.90	0.3432
MCT_RAW	-1.1606427	0.7012105	2.74	0.0979
MST_RAW	-1.6817885	0.7468028	5.07	0.0243*
RCT_RAW	0.51841513	0.633603	0.67	0.4132
SAT_RAW	0.13949123	0.5626549	0.06	0.8042

Fit Y by X Group

Logistic Fit of API_Status By ANI_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	17.91351	1	35.82703	<.0001*
Full	338.81344			
Reduced	356.72696			

RSquare (U)	0.0502
Observations (or Sum Wgts)	3093

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.1328116	0.129625	584.11	<.0001*
ANI_RAW	-1.341877	0.233727	32.96	<.0001*

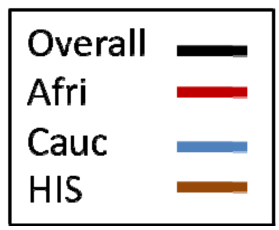
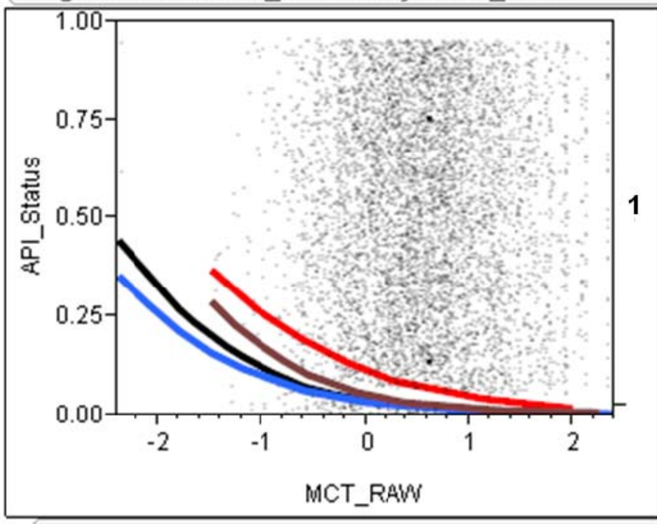


Figure 17. Group API Model Graph for ANIT

Logistic Fit of API_Status By MCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	23.57871	1	47.15743	<.0001*
Full	333.14824			
Reduced	356.72696			

RSquare (U)	0.0661
Observations (or Sum Wgts)	3093

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.3176294	0.1178363	792.68	<.0001*
MCT_RAW	-1.2863815	0.1935332	44.18	<.0001*

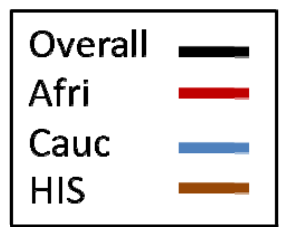
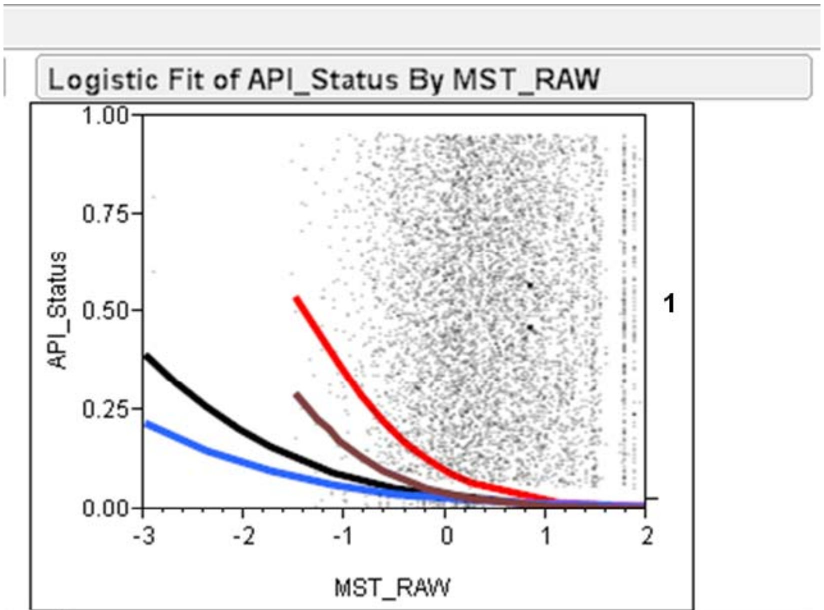


Figure 18. Group API Model Graph for MCT



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	15.78336	1	31.56672	<.0001*
Full	340.94359			
Reduced	356.72696			

RSquare (U) 0.0442

Observations (or Sum Wgts) 3093

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.4556322	0.116996	872.39	<.0001*
MST_RAW	-1.0027598	0.1823302	30.25	<.0001*

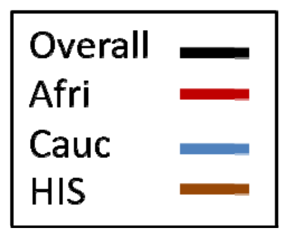
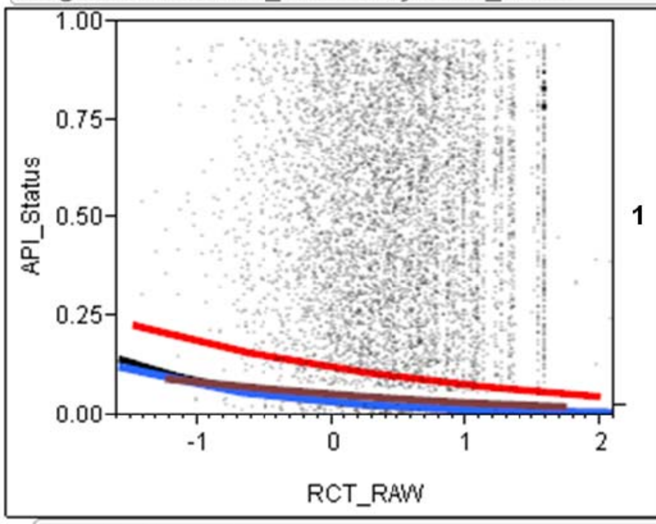


Figure 19. Group API Model Graph for MST

Logistic Fit of API_Status By RCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.94337	1	19.88674	<.0001*
Full	346.78359			
Reduced	356.72696			

RSquare (U)	0.0279
Observations (or Sum Wgts)	3093

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.352098	0.1252476	716.30	<.0001*
RCT_RAW	-0.9565605	0.2160079	19.61	<.0001*

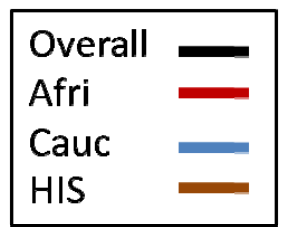
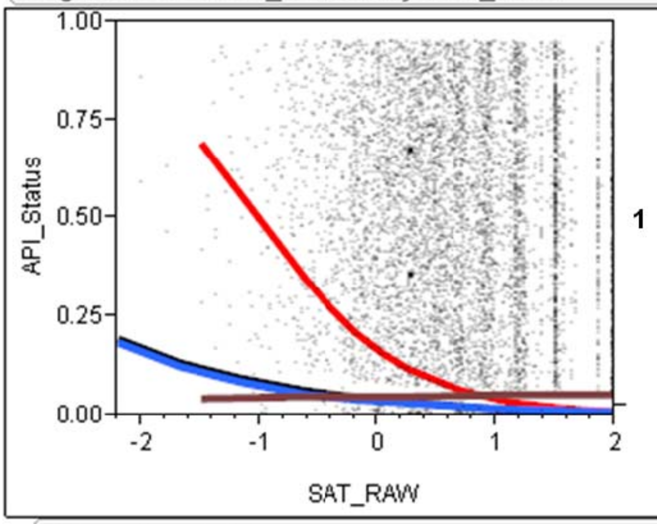


Figure 20. Group API Model Graph for RCT

Logistic Fit of API_Status By SAT_RAW

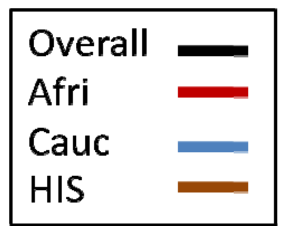


Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	11.82485	1	23.6497	<.0001*
Full	344.90210			
Reduced	356.72696			

RSquare (U)	0.0331
Observations (or Sum Wgts)	3093

Converged by Gradient



Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[1]	-3.2902002	0.1268532	672.73	<.0001*
SAT_RAW	-0.834172	0.1719233	23.54	<.0001*

Figure 21. Group API Model Graph for SAT

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. IFS LOGISTIC REGRESSION MODEL RESULTS AND GRAPHS

Table 20. Overall IFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.27854	5	18.55708	0.0023*
Full	567.44412			
Reduced	576.72266			

RSquare (U)	0.0161
Observations (or Sum Wgts)	3239

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.6873206	0.1383335	377.38	<.0001*
ANIT_RAW	-0.1418997	0.1868099	0.58	0.4475
MCT_RAW	-0.3214514	0.1539625	4.36	0.0368*
MST_RAW	-0.1777812	0.1496344	1.41	0.2348
RCT_RAW	-0.0161861	0.1722276	0.01	0.9251
SAT_RAW	-0.2679942	0.1349024	3.95	0.0470*

Table 21. African American IFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	0.767067	5	1.534134	0.9091
Full	15.641964			
Reduced	16.409031			

RSquare (U)	0.0467
Observations (or Sum Wgts)	91

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-3.1028067	0.7597009	16.68	<.0001*
ANIT_RAW	-0.4541572	1.0478989	0.19	0.6647
MCT_RAW	-0.4414755	1.0665468	0.17	0.6789
MST_RAW	0.39701985	1.0752871	0.14	0.7120
RCT_RAW	0.8750127	1.1624887	0.57	0.4516
SAT_RAW	-0.3128549	0.8702765	0.13	0.7192

Table 22. Caucasian IFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	10.71531	5	21.43061	0.0007*
Full	505.01003			
Reduced	515.72533			

RSquare (U)	0.0208
Observations (or Sum Wgts)	2910

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.6404754	0.1492792	312.87	<.0001*
ANIT_RAW	-0.098538	0.1988217	0.25	0.6202
MCT_RAW	-0.4279137	0.1626452	6.92	0.0085*
MST_RAW	-0.1098234	0.1576678	0.49	0.4861
RCT_RAW	-0.0718655	0.184765	0.15	0.6973
SAT_RAW	-0.3101552	0.142369	4.75	0.0294*

Table 23. Hispanic IFS Logistic Regression Model Results

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2.984260	5	5.968519	0.3093
Full	41.575659			
Reduced	44.559918			

RSquare (U)	0.0670
Observations (or Sum Wgts)	238

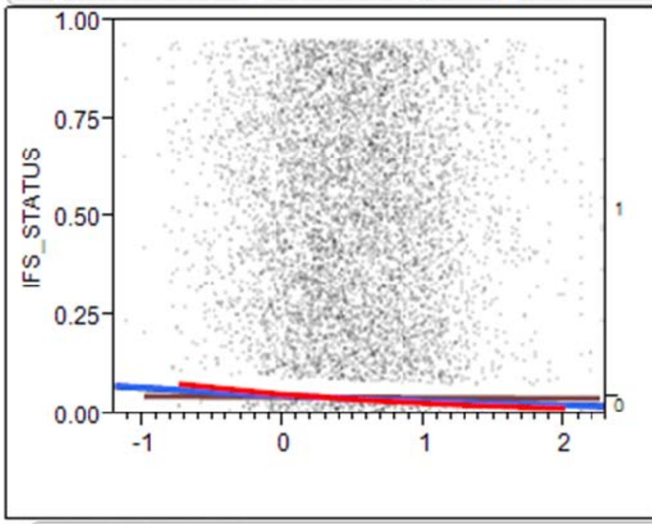
Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-3.2206603	0.5196613	38.41	<.0001*
ANIT_RAW	-0.3734624	0.709614	0.28	0.5987
MCT_RAW	0.89609756	0.5816124	2.37	0.1234
MST_RAW	-1.2777992	0.621506	4.23	0.0398*
RCT_RAW	0.09126423	0.6273483	0.02	0.8843
SAT_RAW	0.02221725	0.4881973	0.00	0.9637

Fit Y by X Group

Logistic Fit of IFS_STATUS By ANIT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1.94703	1	3.894056	0.0485*
Full	574.77564			
Reduced	576.72266			

RSquare (U)	0.0034
Observations (or Sum Wgts)	3239

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.9432712	0.112308	686.81	<.0001*
ANIT_RAW	-0.3417155	0.1741696	3.85	0.0498*

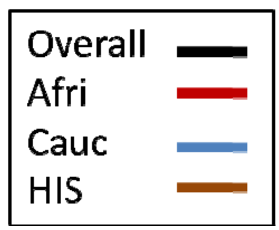
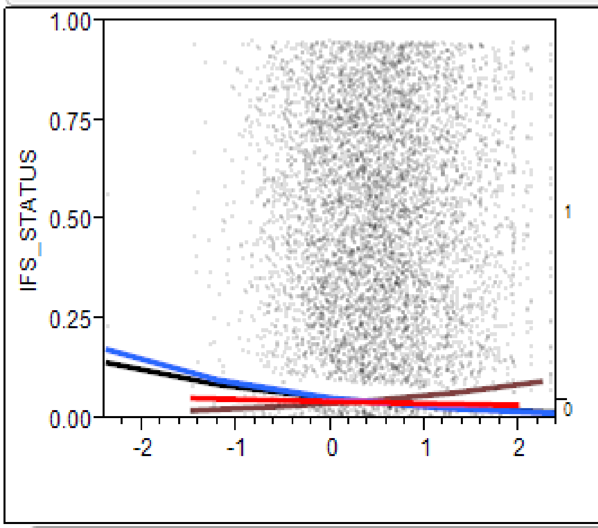


Figure 22. Group IFS Model for ANIT

Logistic Fit of IFS_STATUS By MCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.08253	1	12.16506	0.0005*
Full	570.64013			
Reduced	576.72266			

RSquare (U)	0.0105
Observations (or Sum Wgts)	3239

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.9097801	0.0963457	912.13	<.0001*
MCT_RAW	-0.4665121	0.1350584	11.93	0.0006*

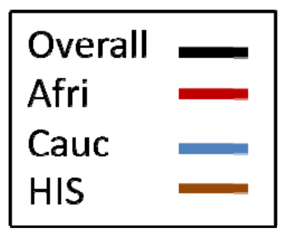
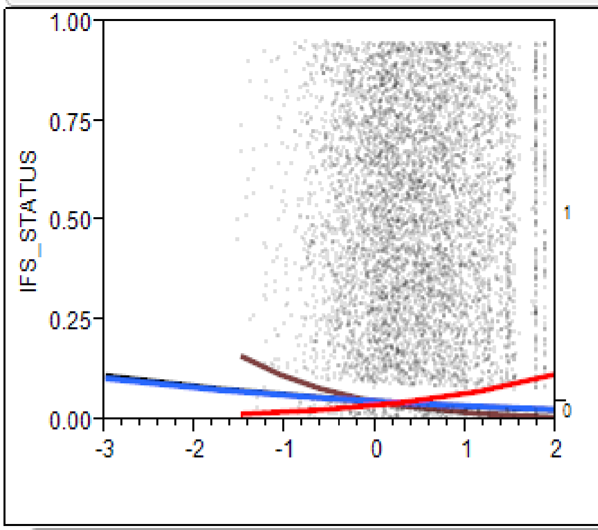


Figure 23. Group IFS Model for MCT

Logistic Fit of IFS_STATUS By MST_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2.60396	1	5.207929	0.0225*
Full	574.11870			
Reduced	576.72266			

RSquare (U)	0.0045
Observations (or Sum Wgts)	3239

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.9780648	0.0973121	936.56	<.0001*
MST_RAW	-0.3059256	0.1346799	5.16	0.0231*

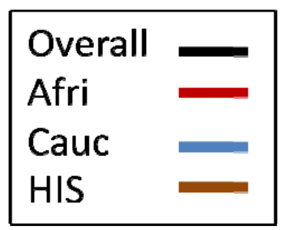
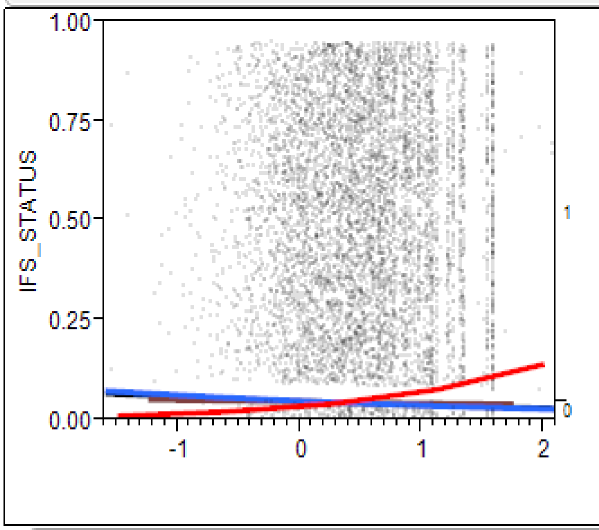


Figure 24. Group IFS Model for MST

Logistic Fit of IFS_STATUS By RCT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	1.02299	1	2.045973	0.1526
Full	575.69968			
Reduced	576.72266			

RSquare (U)	0.0018
Observations (or Sum Wgts)	3239

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.991804	0.110482	733.30	<.0001*
RCT_RAW	-0.2294455	0.1604388	2.05	0.1527

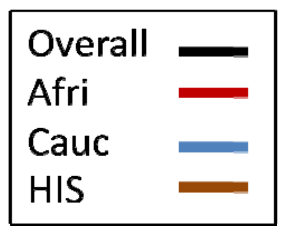
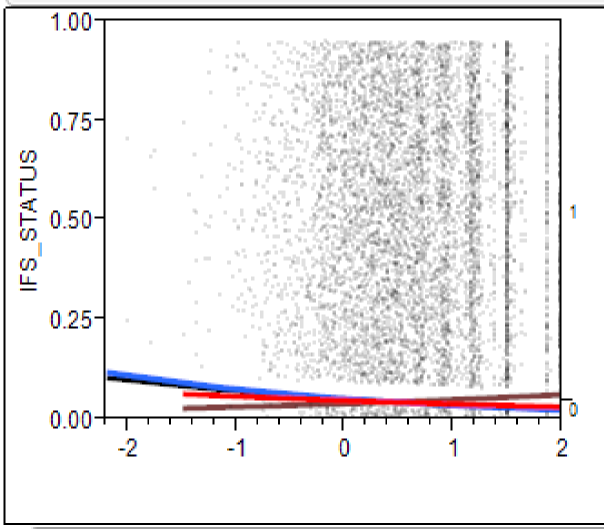


Figure 25. Group IFS Model for RCT

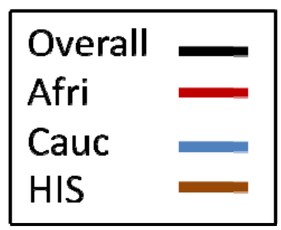
Logistic Fit of IFS_STATUS By SAT_RAW



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	3.76814	1	7.536284	0.0060*
Full	572.95452			
Reduced	576.72266			

RSquare (U)	0.0065
Observations (or Sum Wgts)	3239



Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-2.9108125	0.1044568	776.52	<.0001*
SAT_RAW	-0.3531974	0.1286057	7.54	0.0060*

Figure 26. Group IFS Model for SAT

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Aamodt, M. G. (2004). *Applied Industrial Organizational Psychology* (4th ed). Belmont, CA: Wadsworth Publishing Co.

ASVAB Official Site (2011). History of military testing. Retrieved March 18, 2011, from, http://www.official-asvab.com/history_coun.htm

Berkshire, J. R. (1967). *Evaluation of Several Experimental Aviation Selection Tests*. Pensacola, FL: Naval Aerospace Medical Institution.

Booher, H. R. (2003). Introduction to human systems integration. *Handbook of Human Systems Integration*. Hoboken, NJ: Wiley-Interscience.

Boyd, A. E. (2003). Analysis of determinants of student pilot success for United States naval academy graduates. Master's thesis, Naval Postgraduate School, Monterey, California.

Burke, E. F., & Hunter, D. R. (1995). *Handbook of Pilot Selection.* Aldershot, England: Ashgate Publishing Limited.

Cascio, W. (1998). *Applied Psychology*. Upper Saddle River, NJ: Simon and Schuster.

Carretta, T. R. (1997). Group differences on US Air Force pilot selection tests. *International Journal of Selection and Assessment*. 115–127. Malden, MA: Blackwell Publishers Ltd.

Carretta, T. R., & Ree, M. J., (2003). Pilot selection methods. In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds), *Principles and Practice of Aviation* Psychology. 357–396. Mahwah, N.J: Lawrence Erlbaum Associates, Inc.

Chief of Naval Air Training Instruction. (2007). CNATRAINST 1500.4G *Student Naval Aviator Training and Administration Manual*. Corpus Christi, TX, Author.

Dean, B. J. (1996). *Aviation selection testing: The effect of minimum scores on minorities*. Master's thesis, Naval Postgraduate School, Monterey, California.

Department of the Army (2005). *Selection and training of army aviation officers (Army Regulation 611–110).* Washington, DC: Government Printing Office.

Department of the Navy, (2010). *Annual report on diversity*. Washington, DC: Government Printing Office.

Fiske, D. W. (1947). Validation of naval aviation cadet selection tests against training criteria. *Journal of Applied Psychology*, 5, 601–614.

Flanagan, J. C. (1942). The selection and classification program for aviation cadets (aircrew-bombardiers, pilots, and navigators). *Journal of Consulting Psychology*, *6*, 22–239.

Gonzalez, M. (2003). *Predictors of aviation service selection among U.S. Naval academy graduates*. (Master's thesis, Naval Postgraduate School. Monterey, California).

Government Accountability Office, (1999). Actions needed to better define pilot requirements and promote retention. Washington, DC: Government Printing Office.

Henmon, V. A. C. (1919). Air service tests of aptitude for flying. *Journal of Applied Psychology*. 2, 103–109.

Hilton, T. F., & Dolgin, D. L. (1991). Pilot selection in the military of the free world. *Handbook of Military Psychology*. New York, NY: Wiley. 81–101.

Jenkins, J. G. (1946). Naval aviation psychology (II): The procurement and selection organization. *American Psychologist*. 1, 45–49.

Katz, L. (2006). Finding the right stuff. *U.S. Army Research Institute.* Fort Rucker, AL: U.S. Army.

Martinseen, M., & Hunter, D. R. (2010). *Aviation Psychology and Human Factors*. Boca Raton, FL: CRC Press.

Navy Aerospace Medical Institute (1991). Chapter 12: Aerospace psychological qualifications. *U. S. Naval Flight Surgeon's Manual*, (3).

Navy Aerospace Medical Institute, (2011). ASTB overview. Pensacola, FL. Retrieved March 16, 2011 from: http://www.med.navy.mil/sites/navmedmpte/nomi/nami/Pages/ASTBOverview.aspx

Norman, G. R., & Streiner, D. L. (2003). *Pretty darned quick statistics*. Hamilton, Ontario: BC Decker.

Olde, B. A., Olson, T. M., & Philips, H. L. (2007, March). *Online Delivery of the Navy Aviation Selection Test Battery (ASTB): Development of the Automated Pilot Examination (APEX) System*. Paper presented at the meeting of Human Systems Integration symposium, Alexandria, Virginia.

Olde, B. A., Olson, T. M., & Walker, P. B. (2007, March). *Improving Aviator Selection Using the Performance-Based Measure Measurement Battery (PBMB).* Paper presented at the meeting of Human Systems Integration symposium, Alexandria, Virginia.

Ostoin, S. D. (2007). *An assessment of the performance-based measurement battery (PBMB), the Navy's psychomotor supplement to the aviation selection test battery (ASTB).* Master's Thesis, Naval Postgraduate School, Monterey, California.

Outzz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance*. Vol 15, 161–171.

Pohlman, D. L., & Fletcher, J. D. (1999). Aviation Personnel Selection and Training in D.J. Garland, J. A. Wise, & V. D. Hopkin (eds). *Handbook of Aviation Human Factors*. Mahwah, NJ Lawrence Erlbaum Associates.

Proctor, R. W. & Van Zandt, T. (2008). Human Factors in Simple and Complex Systems (2nd ed). Boca Raton, FL: CRC Press.

Reinhart, P. M. (1998). *Determinants of flight training performance: Naval Academy classes of 1995 and 1996*. Master's thesis, Naval Postgraduate School, Monterey, California.

Roughead, G. (2009). *Statement of Chief of Naval Operations before the House Armed Services Committee on FY10 Department of Navy Posture.*

SAS Institute Inc. (2010). JMP 9 Basic Analysis and Graphing. Cary, NC.

Smart, M. (2008). *Learn Excel 2007 Essential Skills with the Smart Method: Courseware Tutorial for Self-Instruction to Beginner and Intermediate Level.* Smart Method Ltd.

U. S. Air Force ROTC, (2009). *Qualifying Test*. Retrieved March 17, 2011, from http://www.afrotc.com/admissions/qualifying-test/

U. S. Army Air Force in World War II (1945). Aircraft accidents—number and rate: fiscal years 1921 to 1945. *Army Air Forces Statistical Digest, World War II.* Retrieved March 18, 2011, from, http://www.usaaf.net/digest/t212.htm

Wahl, E. J. (1998). *An analysis of aviation test scores to characterize student naval aviator disqualification*. Master's thesis, Naval Postgraduate School, Monterey, California.

Wiener, S. (2005). *Military Flight Aptitude Tests.* 6th Ed, 38 Lawrenceville, NJ.

Williams, H. P., Albert A.O., & Blower D. J., (2000). *Selection of Officers for US Naval Aviation Training, Pensacola, Florida*: Naval Aerospace Medical Research Lab.

Yerkes, R. L. (1921). Psychological examining in the U.S. Army. *Memoirs of the National Academy of Sciences*. (Vol 15). Washington, DC: National Academy of Sciences.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. LCDR Hank Philips, USN
   Naval Operational Medicine Institute
   Pensacola, Florida

4. LDCR Chris Foster, USN
   Chief, Naval Education and Training
   Corpus Christi, Texas